Univerzita Karlova v Praze
Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

Dizertační práce

Mgr. Petr Klášterecký

Odhady parametrů v subkohortních studiích

Školitel: Mgr. Michal Kulich, Ph.D.
Studijní obor: M4 – Pravděpodobnost a matematická statistika

Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Doctoral Thesis

Mgr. Petr Klášterecký

Parameter estimation in case-cohort studies

Supervisor: Mgr. Michal Kulich, Ph.D.
Branch: M4 – Probability and mathematical statistics

Acknowledgement

I am grateful to my advisor Mgr. Michal Kulich, Ph.D. for cooperation, consultations and numerous suggestions that greatly enhanced the presentation. I am also grateful to my family, colleagues and friends for their continual support and encouragement.

The research was partially supported by the Czech national science foundation grant GAČR 201/05/H007.

Prague, September 14, 2009

Abstrakt

Název práce: Odhady parametrů v subkohortních studiích

Autor: Mgr. Petr Klášterecký

Katedra: Katedra pravděpodobnosti a matematické statistiky

Školitel: Mgr. Michal Kulich, Ph.D.

E-mail školitele: kulich@karlin.mff.cuni.cz

Abstrakt:

Dizertační práce se zabývá odhadem parametrů v regresních modelech v analýze přežívání, zejména v tzv. subkohortních studiích. V těchto studiích se provádí výběr pozorování do subkohorty, která se sleduje a analyzuje. To umožňuje výrazné snížení nákladů na provedení studie, ale zároveň vyžaduje při odhadu regresních parametrů jiné postupy než klasická analýza přežívání. Obvyklá úprava odhadovacích rovnic spočívá v zavedení váhových funkcí, které zohledňují pravděpodobnost výběru jednotlivých pozorování do subkohorty. V práci ukážeme, že tato metoda může vést při malých pravděpodobnostech výběru k vychýleným odhadům, a navrhneme alternativní odhad založený na logistické regresi.

Klíčová slova:

Subkohortní studie, analýza přežívání, logistická regrese, vážený odhad.

Abstract

Title: Parameter estimation in case-cohort studies

Author: Mgr. Petr Klášterecký

Department: Dept. of Probability and Mathematical Statistics

Supervisor: Mgr. Michal Kulich, Ph.D.

Supervisor's e-mail address: kulich@karlin.mff.cuni.cz

Abstract:

The concern of this thesis is parameter estimation in regression models in survival analysis, particularly in case-cohort studies. In case-cohort studies, observations are sampled to form a subcohort which is followed and analysed. As a result, the cost of performing such studies is reduced but standard procedures for parameter estimation need to be modified. This is usually done by incorporating weights into the estimating equations so that individual sampling probabilities are accounted for. In this thesis we show that this method can lead to biased estimators when the subcohort sampling probability is low and suggest an alternative estimator based on logistic regression.

Keywords:

Case-cohort study, survival analysis, logistic regression, weighted estimator.

Contents

1 Introduction **1**

2 Regression models for survival data analysis **3**
2.1 Basic concepts, censoring, notation 3
2.2 Proportional hazards model 5
2.3 Proportional odds model 7
2.4 Nonlinear transformation models 10
2.5 Logistic regression model 13

3 Parameter estimation under the case-cohort design **15**
3.1 Introduction to the case-cohort design 15
3.2 Fitting the proportional hazards model to case-cohort data . . 17
3.3 Other regression models under the case-cohort design 21
3.4 Performance of case-cohort estimators 26
3.5 Logistic models and case-control data 31

4 The Combined Logistic Estimator **37**
4.1 Background 37
4.2 Construction of the estimator 42
4.3 Theoretical results 46

5 A numerical study **65**
5.1 Simulation design 65
5.2 Results 66

6 Summary and discussion **75**

References **79**

Chapter 1

Introduction

Survival analysis as a standalone statistical discipline emerged more than 40 years ago when the basic methods for analysing censored and/or truncated time to event data were developed. The original Kaplan-Meier estimator for the survival function (developed around 1960) and the Cox proportional hazards model (1972) became extremely popular and are still widely applied in practice. On the other hand, survival analysis remains an area of active research and many papers suggesting new approaches, models and estimation methods recently appeared in the literature.

Similarly the original paper on the *case-cohort design* by R.L. Prentice was published long ago in 1986 and new papers appear on a regular basis. The case-cohort design aims to reduce costs of the study by observing much fewer subjects compared to classical study designs. Rather than following the whole cohort, a random sample called *subcohort* of the individuals is selected and then followed. As a result, case-cohort studies still generate time to event data, which can be analysed by known models. Traditional estimation methods taken from survival analysis cannot be directly used for the analysis, most usually they are modified in order to reflect the sampling scheme. Nevertheless, the extensive publishing on survival analysis had a strong impact on research of case-cohort studies and parameter estimation in case-cohort regression models.

Regression models for time to event data and parameter estimation under the case-cohort design are the main topics of this doctoral thesis. Since there are many regression models for time to event data, many authors aimed to unify the models and estimating procedures. The resulting class of nonlinear transformation models will be introduced in Chapter 2, however most attention will be paid to the Cox proportional hazards model and the proportional odds model. While the Cox model is well known and popular, the latter model often seems to be a reasonable but neglected alternative.

All of the most important regression models known from survival analysis can be used for analysing case-cohort data as well. Many case-cohort estimators have been suggested throughout the literature and their development closely followed the development seen in survival analysis. Originally the estimators have been developed for specific models, e.g. the Prentice's estimator from 1986 deals with the Cox proportional hazards model. Later there were efforts to unify the treatment and to modify or generalise estimating equations to case-cohort data in the whole class of nonlinear transformation models. The main ideas of parameter estimation in case-cohort studies are presented in Chapter 3.

Currently used case-cohort estimators introduce some weight functions into the estimating equations. They are based on the same principle and therefore they also share similar weaknesses in situations where the probability of observing an event is generally very low regardless of the covariates. In simulation studies we have often seen bias and inaccurate coverage of 95% confidence intervals. A discussion on reasons for such performance problems and a small simulation study illustrating this behaviour are also reported in Chapter 3. These issues are very interesting for practical use of case-cohort estimators, since situations with low event rates occur frequently in practice and the case-cohort design would save most resources here.

Most of the original results are presented in Chapter 4, where we develop a new estimator for regression parameters in the proportional odds model and its asymptotic properties. The estimating procedure is based on combining logistic regression estimates made in subsequent failure times. To obtain the estimator we choose a convex linear combination of the individual logistic estimators that minimizes the asymptotic variance of each estimated parameter component. Performance of the estimator is then illustrated in a simulation study presented in Chapter 5. Note that our estimator performed very well with data generated from the popular proportional hazards model, although it was developed for the proportional odds model. This behaviour is also theoretically discussed in Chapter 4.

The thesis is concluded with a summary and discussion on open problems in Chapter 6.

Chapter 2

Regression models for survival data analysis

In this chapter we shall introduce several regression models or model families commonly used for analysing survival data. However, before we can proceed to regression models it is necessary to introduce some notation, distributional characteristics frequently used in survival analysis and the concept of censoring. The next section only covers the very basic quantities used throughout the whole work. Additional notation will be introduced in later chapters as needed.

2.1 Basic concepts, censoring, notation

Assume there are a nonnegative continuous random variable T, usually called "failure time" or "event time" and a p-dimensional vector of covariates $\boldsymbol{Z}$. Covariates can generally be time-dependent, however we shall only consider fixed covariates in this work. The vector of covariates is bound to the failure time through an unknown p-dimensional regression parameter $\boldsymbol{\beta}$. Later on we shall use $\boldsymbol{\beta}$ to denote the regression parameter in general while $\boldsymbol{\beta}_0$ will always denote the true value of $\boldsymbol{\beta}$. Denote the conditional distribution function of T by $F_{\boldsymbol{Z}}(t) = P(T \leq t|\boldsymbol{Z})$ and the conditional survival function $S_{\boldsymbol{Z}}(t) = 1 - F_{\boldsymbol{Z}}(t) = P(T > t|\boldsymbol{Z})$. In situations where confusion might occur we will explicitly state the corresponding random variable in the lower index, e.g. $F_{T|\boldsymbol{Z}}(t)$.

The conditional hazard rate defined by (2.1) is another important characteristic of the failure time distribution:

$$\lambda(t|\boldsymbol{Z}) = \lim_{h \searrow 0} \frac{1}{h} P[t \leq T < t + h|T \geq t, \boldsymbol{Z}]. \tag{2.1}$$

The conditional hazard rate is often viewed as the instantaneous probability of failure at $T = t$ given that $T \geq t$. It is directly related to the conditional survival function since clearly

$$\lambda(t|\boldsymbol{Z}) = \frac{\partial}{\partial t}\left[-\log(S(t|\boldsymbol{Z}))\right].$$

Modelling the conditional survival function or its transformations is essential for the so called nonlinear transformation models, see Sections 2.3 and 2.4. On the other hand, the conditional hazard function plays a major role in the proportional hazards model, see Section 2.2. Regardless of exploiting survival or hazard functions for modelling there are two features common to all regression models considered in this work. All the models are semiparametric (meaning they contain real-valued as well as functional parameters) and the dependence of the event time T on $\boldsymbol{Z}$ will be always modelled through a linear term $\boldsymbol{\beta}'\boldsymbol{Z} = \beta_1 Z_1 + \ldots + \beta_p Z_p$. Linearity of the regression predictor has a clear advantage of simplicity and the regression parameters $\beta_1, \ldots, \beta_p$ often have a practically useful interpretation.

Right censoring

Censoring (in any form) is the most important issue that distinguishes survival analysis from other parts of mathematical statistics. We will restrict our attention to censoring on the right since it is the most natural and most often arising censoring pattern in epidemiological and biological applications. More details on left or interval censoring can be found in the literature (see e.g. Kalbfleisch & Prentice, 2002, Chap. 1 and 3). The most important consequence of right censoring is that it is impossible to observe exact event times for some objects in a study and this fact has to be taken into account during the analysis of censored data.

Assume that besides the failure time T there is a random variable C called the censoring time. We say that T is *right-censored* if we only observe $X = \min(T, C)$ and the indicator variable $\delta = \mathbb{I}_{[X=T]}$ instead of T itself. The random variable $X = \min(T, C)$ is called *the censored failure time* and δ will be referred to as *the censoring indicator*.

Right censoring occurs naturally when the study ends before all enrolled subjects experienced the event of interest, when subjects are lost during the study due to moving to another location or, indeed, due to reasons directly connected with the study such as worsening or improving of their health condition in medical studies. It is intuitively apparent that some censoring mechanisms will introduce bias to the analysis more likely than others. When the event time further depends on some covariate vector $\boldsymbol{Z} = \boldsymbol{z}$, the "harmless" censoring mechanism can be formalized mathematically through the hazard rate. Roughly speaking, the hazard rate should remain the same whether or not information on censoring is available. This notion is referred to as *independent censoring* in the literature.

Definition 2.1: Independent censoring. *The censoring mechanism is called independent, if*

$$\lim_{h \searrow 0} \frac{P(T \in [t, t+h) | z, T \geq t)}{h} = \lim_{h \searrow 0} \frac{P(T \in [t, t+h) | z, T \geq t, C \geq t)}{h}$$

holds for almost all $t \in \mathbb{R}^+$.

Independent censoring covers many of the common censoring patterns including the so-called Type I and Type II censoring (censoring up to a given time or up to a given number of events) and is most often assumed by standard techniques of survival analysis.

Counting processes

We can see that each study subject in survival analysis can be in several states at a given time t. The subject either has already experienced the event of interest, has been censored prior to t or still remains *at risk* at time t. Such behaviour is conveniently mathematically expressed through stochastic processes defined so that $N(t) = \mathbb{I}_{[T \leq t,\ \delta = 1]}$ is the event counting process and $Y(t) = \mathbb{I}_{[X > t]}$ is the at risk process. Note that $Y(t) = 0$ implies that a potential failure at time t cannot be observed. In the following we will suppose there are n subjects in the study providing n independent realisations of T, C, and $\boldsymbol{Z}$. Note however that due to censoring we are only able to observe triplets $(X_i, \delta_i, \boldsymbol{Z}_i)$ or $(N_i, Y_i, \boldsymbol{Z}_i)$ for $i = 1, \ldots, n$.

2.2 Proportional hazards model

Results concerning the proportional hazards model are usually well known and reviewed in many standard textbooks such as Kalbfleisch & Prentice (2002) or Fleming & Harrington (1991). We shall therefore only point out the basic ideas here and later in the text we shall concentrate on comparison with other regression models in survival analysis.

The proportional hazards model specifies dependence of survival time on covariates through the conditional hazard rate. Suppose that for any two covariate values $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ the associated failure rates have a fixed ratio over time. Then $\lambda(t|\boldsymbol{Z}_1) = k(\boldsymbol{Z}_1, \boldsymbol{Z}_2)\lambda(t|\boldsymbol{Z}_2)$ where k is a nonnegative real valued function which does not depend on time. We say that the hazard rates are *proportional.* If we denote *the baseline hazard rate* by $\lambda_0(t) = \lambda(t|\boldsymbol{Z} = 0)$, then

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)k(\boldsymbol{Z}, \mathbf{0}) = \lambda_0(t)g(\boldsymbol{Z}).$$

The function $g(\boldsymbol{Z})$ must not be negative (modelling a hazard rate) with $g(\mathbf{0}) = 1$ and its dependence on $\boldsymbol{Z}$ should not be too complicated in order to keep the model practically useful. The most common form of the proportional hazards model is therefore

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}), \tag{2.2}$$

which assumes that g depends on $\boldsymbol{Z}$ through the linear combination $\boldsymbol{\beta}_0'\boldsymbol{Z}$.

The unknown parameters in (2.2) are $\boldsymbol{\beta}_0$ and $\lambda_0(t)$. The vector of regression parameters $\boldsymbol{\beta}_0$ is the main parameter of interest while $\lambda_0(t)$ is treated as an infinitely dimensional (functional) nuisance parameter. For $\beta_{0j},\ j = 1, \ldots, p$, the term $e^{\beta_{0j}}$ from (2.2) shows the relative risk of failure for an individual with covariate vector $Z_1, \ldots, Z_{j-1}, Z_j + 1, Z_{j+1}, \ldots, Z_p$ as compared to an individual with covariates $Z_1, \ldots, Z_{j-1}, Z_j, Z_{j+1}, \ldots, Z_p$.

Parameter estimation

Standard likelihood techniques cannot be used for parameter estimation in (2.2) due to the presence of $\lambda_0(t)$. Instead, the concept of partial likelihood, see (Cox, 1972) and (Cox, 1975), is used for estimating the vector of parameters $\boldsymbol{\beta}_0$ from the observed data. The natural logarithm of the partial likelihood is then treated in much the same way as an ordinary log-likelihood function in the sense that statistical inference is based on its first and second derivatives. The main advantage of partial likelihood is that it eliminates the unknown baseline hazard function $\lambda_0(t)$, see Definition 2.2.

Definition 2.2: *The partial likelihood function for the proportional hazards model* (2.2) *is given by*

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{s\geq 0}\left\{\frac{Y_i(s)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)}{\sum_{l=1}^{n} Y_l(s)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}\right\}^{\Delta N_i(s)}. \tag{2.3}$$

The maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}$ can be obtained by setting the partial derivatives of the logarithm of (2.3) with respect to $\boldsymbol{\beta}$ equal to 0 and solving a system of equations

$$\begin{aligned}\mathbf{0} = \boldsymbol{U}(\boldsymbol{\beta}) &= \frac{\partial}{\partial\boldsymbol{\beta}}\log L(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n}\int_0^\infty \{\boldsymbol{Z}_i(u) - \mathsf{E}(\boldsymbol{\beta}, u)\}\, dN_i(u),\end{aligned} \tag{2.4}$$

where

$$\mathsf{E}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} = \frac{n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i Y_i(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{n^{-1} \sum_{i=1}^{n} Y_i(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}. \tag{2.5}$$

This way one can view the score vector as a comparison of the observed values of covariates of the failed individual and their expected values for individuals remaining at risk. Using martingale methods one can prove (under certain regularity conditions) consistency of the partial maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ and weak convergence of the normalized score process to a Gaussian stochastic process. The asymptotic results are therefore similar to those from the classical maximum likelihood estimation. For more details we refer to standard textbooks such as (Fleming & Harrington, 1991, Chap. 4).

2.3 Proportional odds model

The proportional odds model was introduced in Bennett (1983) as a generalisation of results derived by McCullagh (1980) to right censored data. The main purpose of Bennett's work was to make McCullagh's results suitable for use in medicine and epidemiology. In the original 1983 paper Bennett proposed to use the proportional odds model as an alternative to Cox's regression in situations when one needs to demonstrate an effective cure, mathematically expressed as hazard ratios converging to unity over time. We will examine this in more detail later in this section.

The proportional odds models is specified through the conditional survival function as

$$-\mathrm{logit}(S(t|\boldsymbol{Z})) = G(t) + \boldsymbol{\beta}_0' \boldsymbol{Z}, \tag{2.6}$$

where $\mathrm{logit}(x) = \log(x/(1-x))$. The unknown parameters are $G(t) = -\mathrm{logit}(S(t|\boldsymbol{Z} = \mathbf{0}))$, a baseline log-odds of failure by time t, and the vector of regression coefficients $\boldsymbol{\beta}_0$. It is convenient and common to work with an alternative parametrization of the model that specifies the baseline odds of failure as $H(t) = \exp(G(t))$. The new functional parameter H is then a nondecreasing right-continuous function with left-hand limits mapping $\mathbb{R}^+$ to $\mathbb{R}^+$ with $H(0) = 0$.

The proportional odds model follows the same semiparametric concept as the proportional hazards model, where the unknown parameters were $\boldsymbol{\beta}_0$ and $\lambda_0(t)$, however the interpretation of the model is naturally different. With continuous covariates, the term $e^{\beta_{0j}}$ here shows the change in the odds of failure per a unit increase in Z_j, keeping all other covariate values fixed.

To illustrate the difference between the proportional hazards and proportional odds models on a simple example suppose for a moment that there is only one binary covariate Z. Such setting brings up a simple two-sample problem representing e.g. patients treated with medication and placebo. The proportional odds model forces the ratio of the odds of survival (or failure) to be constant over time, more specifically

$$\frac{S(t|Z=1)}{1-S(t|Z=1)} : \frac{S(t|Z=0)}{1-S(t|Z=0)} = \frac{\exp(-G(t)-\beta)}{\exp(-G(t))} = \exp(-\beta)$$

yielding

$$\frac{S(t|Z=1)}{1-S(t|Z=1)} = \frac{S(t|Z=0)}{1-S(t|Z=0)}\exp(-\beta).$$

Differentiating logarithms of both sides with respect to t gives

$$\frac{\partial \log(S(t|Z=1))}{\partial t} - \frac{\partial \log(1-S(t|Z=1))}{\partial t} \\ = \frac{\partial \log(S(t|Z=0))}{\partial t} - \frac{\partial \log(1-S(t|Z=0))}{\partial t}$$

which can be simplified to

$$\lambda(t|Z=1) - \lambda(t|Z=0) = \frac{f(t|Z=1)}{F(t|Z=1)} - \frac{f(t|Z=0)}{F(t|Z=0)}. \tag{2.7}$$

As $t \to \infty$, the right hand side of (2.7) converges to 0 and thus $\lambda(t|Z=1)$ converges to $\lambda(t|Z=0)$ in the proportional odds model. In the proportional hazards model, however, the ratio of hazards is held constant.

Despite this difference, there are situations where both models provide similar conclusions. This occurs when the overall probability of the event of interest is low regardless of covariate values. With one binary covariate it means that $S(t|Z=1) \approx S(t|Z=0) \approx 0$ and

$$\frac{S(t|Z=1)}{1-S(t|Z=1)} : \frac{S(t|Z=0)}{1-S(t|Z=0)} \approx \frac{S(t|Z=1)}{S(t|Z=0)}.$$

The relative risk is thus approximated with the odds ratio. The possibility of using the proportional odds model as an approximation to the proportional hazards model when working with rare events generates new application opportunities for the proportional odds model.

Parameter estimation

In the proportional hazards model (2.2) it was convenient to use the partial likelihood (2.3) for estimation of $\boldsymbol{\beta}$, because it eliminated the problem with estimating the unknown baseline hazard function. There is no such elegant solution in the case of the proportional odds model. Bennett (1983) proposed transforming the the failure times to the log-logistic distribution and treat such transformed failure times as nuisance parameters while estimating $\boldsymbol{\beta}$. This approach is known as using a profile likelihood, the functional part G is profiled out. Since there may be almost as many parameters to estimate as there are observations, this method can lead to biased results, especially in small samples.

The original Bennett's paper contains no results regarding asymptotic properties of his estimator. These were established later by Murphy et al. (1997). Murphy et al. worked with a slightly reparametrized model using the baseline odds of failure $H(t) = \exp(G(t))$. They showed that the original Bennett's estimator is consistent and asymptotically normal with an efficient variance and that the estimator of $H(t)$, which is a nondecreasing step function with jumps in the observed failure times, is uniformly consistent.

The contribution to the likelihood for one observation in a right-censored dataset was shown by Murphy et al. to be

$$\begin{aligned}\mathrm{L}(X, \delta, \boldsymbol{Z}, H, \boldsymbol{\beta}) = &\left(\frac{e^{-\boldsymbol{Z}'\boldsymbol{\beta}}}{(H(X) + e^{-\boldsymbol{Z}'\boldsymbol{\beta}})(H(X-) + e^{-\boldsymbol{Z}'\boldsymbol{\beta}})}\Delta H(X)\right)^{\delta} \\ &\times \left(\frac{e^{-\boldsymbol{Z}'\boldsymbol{\beta}}}{H(X) + e^{-\boldsymbol{Z}'\boldsymbol{\beta}}}\right)^{1-\delta},\end{aligned} \tag{2.8}$$

where $X = \min(T, C)$ as introduced earlier and $\Delta H(X) = H(X) - H(X-)$. The profile log-likelihood for $\boldsymbol{\beta}$ is then given by

$$\mathrm{PrL}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log(\mathrm{L}(X_i, \delta_i, \boldsymbol{Z}_i, \hat{H}_{\boldsymbol{\beta}}, \boldsymbol{\beta})), \tag{2.9}$$

where $\hat{H}_{\boldsymbol{\beta}}$ maximizes the log-likelihood for a fixed $\boldsymbol{\beta}$. The maximum profile likelihood estimator $\hat{\boldsymbol{\beta}}$ then maximizes (2.9).

It was also shown that differentiation of the profile log-likelihood yields consistent estimators of the information matrix and that the profile likelihood ratio statistics can be compared with χ^2 percentiles to produce asymptotic tests just like an ordinary likelihood function.

2.4 Nonlinear transformation models

By "nonlinear transformation models" we understand a broad class of semiparametric models where an unknown or unspecified transformation of survival time is linearly related to the vector of covariates. A nonlinear transformation model is thus a regression model specified through the equation

$$H(T) = -\boldsymbol{\beta}_0' \boldsymbol{Z} + \varepsilon, \tag{2.10}$$

where T is the failure time and H is an unknown monotone transformation function such that $H(0) = -\infty$. The error term ε is a random variable with a known and completely specified distribution function F_ε and is assumed to be independent of the covariate vector $\boldsymbol{Z}$.

The nonlinear transformation model (2.10) can be alternatively described by the equation

$$g(S_{\boldsymbol{Z}}(T)) = H(T) + \boldsymbol{\beta}' \boldsymbol{Z}, \tag{2.11}$$

where g is a known decreasing function such that $F_\varepsilon(\cdot) = 1 - g^{-1}(\cdot)$ and the remaining terms have the same meaning as above in (2.10). This is obvious since F_ε is the distribution function of ε.

This general class contains many regression models, including the proportional hazards and proportional odds models, as special cases. Representation of a particular model and therefore also interpretation of the regression parameters are a matter of choice of an appropriate error term distribution. Using the extreme value distribution for ε we obtain the proportional hazards model while the proportional odds model arises with ε distributed as a standard logistic random variable.

To see this use (2.11) first with ε following the extreme value distribution with $F_\varepsilon(x) = 1 - \exp(-\exp(x))$. By substitution we obtain the link $g(\cdot) = \log(-\log(\cdot))$ so that (2.11) becomes $\log(-\log(S_{\boldsymbol{Z}}(t))) = H(t) + \boldsymbol{\beta}'\boldsymbol{Z}$, which is the proportional hazards model expressed in terms of the survival function. Now take ε distributed according to the standard logistic distribution. It follows then that the transformation function $g(\cdot) = (1 - F)^{-1}(\cdot)$ must have the form $g(\cdot) = -\text{logit}(\cdot)$ and (2.11) becomes the proportional odds model (2.6).

As a consequence, any estimation method developed for the general class of the nonlinear transformation models can be used with either of those two widely used regression models. Some of the estimating equations presented later in this section simplify to the partial likelihood score equations in the case of the proportional hazards model, however, there is no such result for the proportional odds model. The original Bennett's profile likelihood estimator (Bennett (1983)) has been proved to be efficient in this particular case. See Murphy et al. (1997) for the proof and further details.

Parameter estimation

Several methods, which mainly differ in their assumptions imposed on the censoring mechanism, have been developed for parameter estimation in nonlinear transformation models. Examples include the method of Cheng et al. (1995) based on the Kaplan-Meier estimator of the "survival" function for the censoring variable C or the method of Chen et al. (2002) with estimating equations motivated by a counting process representation of the model.

We shall first outline the approach of Cheng et al. (1995) here without going into much detail. The paper is mostly interesting because it is probably the first work developing a unified estimating procedure for nonlinear transformation models and because there were many attempts to generalize it to case-cohort data. However, since it has been published, new methods with less restrictive assumptions appeared in the literature.

Cheng et al. assume independence of time to event T and the censoring variable C and also independence of C and covariates $\boldsymbol{Z}$ (although this assumption was shown not to matter for discrete covariates). Using these two key assumptions, Cheng et al. developed generalized estimating equations for $\boldsymbol{\beta}$ as if there were no censoring at all and then used the Kaplan Meier estimator for G to replace quantities that were unobservable due to censoring. Finally they showed that the modified estimating equations have an asymptotically unique solution $\hat{\boldsymbol{\beta}}$, established the asymptotic normality of $\hat{\boldsymbol{\beta}}$ and calculated its asymptotic variance-covariance matrix.

These results were further extended in Cheng et al. (1997) to predicting the survival function and its quantiles from nonlinear transformation models. However, Fine et al. (1998) noted that the Cheng's estimator is asymptotically biased if the support of the censoring variable is shorter than the support of failure time. They suggested an improvement to both the estimator and confidence limits for the survival function. The approach of Fine et al. served as a basis for one of the case-cohort estimators in nonlinear transformation models, see Section 3.

Although Cheng et al. (1995) report good properties of their estimator based on some simulation studies, their assumptions may be difficult to meet particularly when analysing observational data. An estimator developed under less restrictive conditions, particularly without the independence of censoring variable and covariates, was therefore presented in Chen et al. (2002) and will be described in the remainder of this section.

Denote by $\lambda_\varepsilon(\cdot)$ the known hazard function and by $\Lambda_\varepsilon(t)$ the cumulative hazard function of ε. Let further

$$M(t) = N(t) - \int_0^t Y(s) d\Lambda_\varepsilon\{\boldsymbol{\beta}_0'\boldsymbol{Z} + H_0(s)\}$$

and recall that both $N(t)$ and $Y(t)$ are quite simple counting processes. It follows from standard counting process theory in survival analysis (Fleming & Harrington, 1991, Chapter 1) that $\int_0^t Y(s)d\Lambda_\varepsilon\{\boldsymbol{\beta}_0'\boldsymbol{Z} + H_0(s)\}$ is a compensator for $N(t)$ and therefore $M(t)$ defined in this way is a martingale process with zero expectation. This fact motivated Chen et al. (2002) to develop a unified estimation procedure for model (2.10) based on the estimating equations

$$\sum_{i=1}^{n}\int_0^\infty \boldsymbol{Z}_i[dN_i(t) - Y_i(t)d\Lambda_\varepsilon\{\boldsymbol{\beta}'\boldsymbol{Z}_i + H(t)\}] = 0 \tag{2.12a}$$

$$\sum_{i=1}^{n}[dN_i(t) - Y_i(t)d\Lambda_\varepsilon\{\boldsymbol{\beta}'\boldsymbol{Z}_i + H(t)\}] = 0 \quad (\text{for } t \geq 0). \tag{2.12b}$$

The first equation (2.12a) is an analogue to the partial likelihood score equation (2.4) from the proportional hazards model, equation (2.12b) is necessary for estimating values of the unknown transformation function at the observed failure times. This was not needed in the proportional hazards model, where the unknown baseline hazard function vanished from the partial likelihood. In fact, the above equations simplify to the partial likelihood score equation (2.4) if the cumulative hazard rate $\Lambda_\varepsilon(t) = \exp(t)$ is plugged in according to the proportional hazards model. The method of Chen et al. can thus also be viewed as a generalization of partial likelihood methods to nonlinear transformation models.

Suppose there are K observed distinct failure times $0 < t_1 < \cdots < t_K < \infty$. The iterative algorithm for obtaining the estimates of H and $\boldsymbol{\beta}$ works in four steps as follows:

Step 1: Fix $\boldsymbol{\beta}$ at an initial value $\boldsymbol{\beta}^{(0)}$.

Step 2: Compute $H^{(0)}(t_1), \ldots, H^{(0)}(t_K)$ by solving equations

$$\sum_{i=1}^{n} Y_i(t_k)\Lambda_\varepsilon\{H(t_k) + \boldsymbol{\beta}'\boldsymbol{Z}_i\} = 1 + \sum_{i=1}^{n} Y_i(t_k)\Lambda_\varepsilon\{H(t_k-) + \boldsymbol{\beta}'\boldsymbol{Z}_i\}$$

for $H^{(0)}(t_k)$, $k = 1, \ldots, K$, with $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$. Recall that $\hat{H}$, the estimator of H, is a nondecreasing step function which only jumps at the observed failure times. We set $H^{(0)}(t_1-) = -\infty$, reflecting the fact that $H(0) = -\infty$. The last sum thus vanishes in the first equation.

Step 3: Update the estimate of $\boldsymbol{\beta}$ by solving equation (2.12a) for $\boldsymbol{\beta}$ with H fixed at the $H^{(0)}$ obtained in Step 2.

Step 4: Set $\boldsymbol{\beta}^{(0)}$ to the new value obtained in Step 3 and repeat steps 2 and 3 until convergence is reached.

Although the estimating equations have to be solved iteratively (except for the special case of the proportional hazards model), standard numerical procedures such as the Newton-Raphson algorithm work quite well and reasonably fast. As a further computational simplification Chen et al. proposed an approximation for estimating H based on first order differences. According to simulation studies however the inaccuracy induced in this way overrides the advantage of simplicity.

Chen et al. (2002) showed that the resulting estimator $\hat{H}$ of H_0 is a nondecreasing step function with jumps in the observed failure times and that the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ is consistent and asymptotically normal under suitable regularity conditions, with a closed-form variance.

2.5 Logistic regression model

We conclude this Chapter with a short overview of the logistic regression model. Logistic regression is not a typical tool for analysing time to event data, mainly because the response variable for this model is a binary indicator rather than some (censored) time. Nevertheless, a logistic model can be used with survival data when we only know which observations experienced the event of interest and which were censored. That means we do not use any time information and the binary censoring indicator δ plays the role of the response variable. The logistic regression model can be described by

$$\operatorname{logit}(P(\delta = 1|\boldsymbol{Z})) = \log \frac{P(\delta = 1|\boldsymbol{Z})}{1 - P(\delta = 1|\boldsymbol{Z})}) = \alpha_0 + \boldsymbol{\beta}_0'\boldsymbol{Z}, \tag{2.13}$$

where $\boldsymbol{\beta}_0$ is the main parameter of interest providing information on odds ratios or a change in odds ratios while α_0 is important for estimating the overall probability of failure $P(\delta = 1|\boldsymbol{Z})$. Compared to the proportional odds model (2.6), the only difference is in the intercept term α_0, which is a constant in (2.13) but was an unspecified function in (2.6).

Parameter estimation

Standard likelihood techniques provide parameter estimates and further inference such as standard errors, confidence intervals and significance tests. Given iid observations of $(\delta_i, \boldsymbol{Z}_i)$ we can express the likelihood as

$$L = \prod_{i=1}^{n} [\pi(\boldsymbol{z}_i)]^{\delta_i} [1 - \pi(\boldsymbol{z}_i)]^{1-\delta_i}, \tag{2.14}$$

where $\pi(\boldsymbol{z}_i)$ is the probability of failure given the covariates, i.e.

$$\pi(\boldsymbol{z}_i) = P(\delta_i = 1 | \boldsymbol{Z}_i = \boldsymbol{z}_i) = \frac{\exp(\alpha + \boldsymbol{\beta}\boldsymbol{z}_i)}{1 + \exp(\alpha + \boldsymbol{\beta}\boldsymbol{z}_i)}.$$

Differentiating the logarithm of (2.14) with respect to individual parameters leads directly to the score vector $\boldsymbol{U}(\alpha, \boldsymbol{\beta})$ with components

$$U(\alpha) = \sum_{i=1}^{n} [\delta_i - \pi(\boldsymbol{z}_i)] \tag{2.15a}$$

$$U(\beta_j) = \sum_{i=1}^{n} z_{i,j} [\delta_i - \pi(\boldsymbol{z}_i)], \; j = 1, \ldots, p. \tag{2.15b}$$

Setting (2.15) equal to $\mathbf{0}$ provides score equations which can be solved with standard iterative algorithms. Logistic regression models are implemented in most statistical software packages.

Finally, using the first order Taylor expansion of $\boldsymbol{U}$ around the true parameter $(\alpha_0, \boldsymbol{\beta}_0)'$ it can be shown that the estimator $(\hat{\alpha}, \hat{\boldsymbol{\beta}})'$ is unbiased and asymptotically normal and also its variance can be easily estimated. Detailed calculations can be found in many standard statistical textbooks (see e.g. McCulloch & Searle, 2001, p. 102).

Chapter 3

Parameter estimation under the case-cohort design

All methods introduced in Chapter 2 assume that all data is available for all individuals from the study cohort. In this chapter we present the concept of the case-cohort study design, show how the standard methods need to be adapted under the case-cohort design and, in a small simulation study, we illustrate some problems of current approaches to case-cohort parameter estimation.

3.1 Introduction to the case-cohort design

In classical cohort studies, a group of subjects or *a cohort* is randomly selected from the target population and followed for a given time period. The covariates are measured according to the study design, the failure or censoring time is recorded for each individual and a suitable regression model is fitted to the data to draw conclusions. If the occurrence of the event of interest is low in the population, cohort studies must be very large to ensure a sufficient number of cases. The covariates of interest may be expensive to measure (blood sample analyses, complicated laboratory tests etc.). The main drawback of this type of study is then the cost of measuring covariates on a large number of subjects.

For estimating probabilities (odds) of rare events a cost saving solution to this problem is the case-control design (see e.g. the review by Breslow, 2005). Under the case-control design, cases and controls are sampled separately and the sampling probability is typically much larger for the cases. The study design is retrospective and the data can be analysed by logistic regression. Many attempts to extend the case-control design to time to event data were published in the literature. They are known as a synthetic or nested case-control design, see e.g. Mantel (1973), Prentice & Breslow (1978) or Goldstein & Langholz (1992), a hybrid retrospective design (Kupper et al., 1975) or a case-base design (Miettinen, 1982). Recently the case-cohort design by Prentice (1986) became very popular.

The main idea of the Prentice's case-cohort design is to sample individuals from the full cohort and add all cases as they appear. This way *a subcohort* is formed consisting of the cases and sampled controls. Only these individuals from the subcohort are used for the analysis. This way the cases are always included and their covariate values are recorded at their failure times, while only a relatively small number of the remaining individuals are available. Under the case-cohort design the same sampled individuals are used throughout the whole time of study duration. On the contrary, new controls are sampled at each failure time to form the risk set in the nested case-control design.

The oversampling of cases would lead to biased results during the analysis if not accounted for. In order to consistently estimate the regression parameters, case-cohort data are often analysed using various modifications of the corresponding procedures developed for complete data; these were outlined in Chapter 2. The only information that can be used with case-cohort data are covariate measurements for the controls and cases sampled into the subcohort[1] and maybe some additional characteristics such as age, gender or some database entries for the whole cohort. The key problem is to modify the estimating equations, eliminate anything that is not observed due to the case-cohort design and account properly for the sampling scheme. Currently a typical way of dealing with these issues is to introduce weighting functions or constants. The weights should give zero weight to all subjects not sampled into the study and a positive weight to all sampled cases and controls. This approach often leads to weighting individual contributions to estimating equations by inverse sampling probabilities.

This concept is general and common to all regression models for case-cohort data. We shall demonstrate it in more detail by applying it to the proportional hazards model in the next section.

[1] We would have to be more careful here if we allowed time dependent covariates, since then we would only observe covariate values at failure times and covariate histories for individuals sampled into the subcohort at the beginning of the study.

3.2 Fitting the proportional hazards model to case-cohort data

Recall that parameter estimation in the proportional hazards model with full data is based on the partial likelihood score equation

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\infty \{\boldsymbol{Z}_i - \mathsf{E}(\boldsymbol{\beta}, u)\}\, dN_i(u) = 0,$$

where $\mathsf{E}(\boldsymbol{\beta}, t)$ is defined as

$$\mathsf{E}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^1(\boldsymbol{\beta}, t)}{S^0(\boldsymbol{\beta}, t)} = \frac{n^{-1} \sum\limits_{i=1}^{n} \boldsymbol{Z}_i Y_i(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{n^{-1} \sum\limits_{i=1}^{n} Y_i(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}.$$

Even though only the cases contribute directly to the summation in the partial likelihood score $\boldsymbol{U}(\boldsymbol{\beta})$, the controls' influence is hidden in the at-risk covariate averages $\mathsf{E}(\boldsymbol{\beta}, t)$. The term $\mathsf{E}(\boldsymbol{\beta}, t)$ contains data that is unobserved under the case-cohort design. It is therefore necessary to assure that any unobserved subjects will contribute zero to the summation. This idea motivates the pseudoscore

$$\boldsymbol{U}_C(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i - \mathsf{E}_C(\boldsymbol{\beta}, u)\}\, dN_i(u), \tag{3.1}$$

where the case-cohort at risk covariate average $\mathsf{E}_C(\boldsymbol{\beta}, t)$ is given by

$$\mathsf{E}_C(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}_C^1(\boldsymbol{\beta}, t)}{S_C^0(\boldsymbol{\beta}, t)} = \frac{n^{-1} \sum\limits_{i=1}^{n} \varrho_i(t) \boldsymbol{Z}_i Y_i(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{n^{-1} \sum\limits_{i=1}^{n} \varrho_i(t) Y_i(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)} \tag{3.2}$$

with $\varrho_i(t)$ being some weight functions or processes. The weights $\varrho_i(t)$ are set to zero for subjects with incomplete data. Various proposals for the choice of $\varrho_i(t)$ have been published in the literature leading to different parameter estimators, see e.g. Kulich & Lin (2004) for a comprehensive overview.

Let the selection of a subject into the subcohort be indicated by a binary random variable ξ so that $\xi_i = 1$ if subject i was sampled to the subcohort and $\xi_i = 0$ otherwise. Let $\mathrm{P}(\xi_i = 1) = \alpha > 0$ be the sampling probability. Since we need to eliminate unobserved data, the weights should be zero whenever the observation is censored and the subject is not sampled to the subcohort, that is $\varrho_i = 0$ whenever $\delta_i = \xi_i = 0$. For the remaining individuals the weights usually involve the inverse sampling probability $1/\alpha$. The estimator $\hat{\alpha}$ (e.g. the empirical proportion of sampled subjects) can be used instead of α itself. According to Robins et al. (1994), parameter estimation can be actually more efficient when the estimator $\hat{\alpha}$ is used.

The original Prentice's estimator was derived through a modification of the standard risk set without the emphasis on weighting by the inverse sampling probabilities. Its estimating equation can be however represented in the form of (3.1) with $\varrho(t) = \xi_i/\alpha$ for t less than the failure time T_i and $\varrho(T_i) = 1/\alpha$. Such weights represent the concept of following the sampled subcohort and adding any unsampled cases whenever their failures are observed. Cases and sampled controls are given the same importance and equal weight of $1/\alpha$. Moreover, (3.2) can be rewritten without the weights as sums over cases and sampled controls, since α cancels out in the formula.

With another approach the contributions from cases are weighted by 1 through their entire at-risk period and weighting by inverse sampling probabilities is applied to controls only. The weights take the form $\varrho_i(t) = \delta_i + (1-\delta_i)\xi_i/\alpha$ and can be estimated by some constant estimator $\delta_i + (1-\delta_i)\xi_i/\hat{\alpha}$ or some time-varying estimator $\delta_i + (1-\delta_i)\xi_i/\hat{\alpha}(t)$, where $\sup_t |\hat{\alpha}(t) - \alpha| \to 0$ in probability. The dependence on t is most often expressed through $Y_i(t),\ i = 1, \ldots, n$. An example of an estimator with time-varying weights is the estimator proposed by Borgan et al. (2000), Estimator II, which is obtained by setting $\hat{\alpha}(t) = \sum_{i=1}^n \xi_i(1-\delta_i)Y_i(t)/\sum_{i=1}^n(1-\delta_i)Y_i(t)$ – the proportion of sampled controls among all controls remaining at risk at time t. We shall refer to this estimator as the BII estimator further on.

The estimators, where the cases are sampled with probability one at their failure times only and the subcohort is treated as a sample of all study subjects are referred to as *N-estimators* by Kulich & Lin (2004). The original Prentice's estimator belongs to this class. The latter, where all cases are separated and the subcohort is considered to be sampled from the controls only, are known as *D-estimators.* Both classes of estimators weight contributions to the estimating equations by inverse sampling probabilities.

Asymptotic properties of the estimators

Several regularity conditions need to be satisfied in order to show consistency and asymptotic normality of the case-cohort estimators. These conditions can be naturally divided into two groups – conditions known from standard survival analysis and new conditions because of the subcohort sampling. The former conditions for inference based on partial likelihood are summarized e.g. in Fleming & Harrington (1991), p. 289-90. They mainly involve restriction to a finite time interval $t \in [0, \tau]$ and asymptotic stability of the information matrix and of population averages $\boldsymbol{S}^{\ell}(\boldsymbol{\beta}, t),\ \ell = 0, 1$ (see (2.5)).

Regularity conditions specific to the case-cohort design may vary slightly from estimator to estimator. Basically they assure the asymptotic stability of case-cohort averages $\boldsymbol{S}_C^{\ell}(\boldsymbol{\beta}, t),\ \ell = 0, 1$ (see (3.2)) and restrict the subcohort sampling probability to be bounded away from zero. We shall now adapt the general treatment of (Kulich & Lin, 2004, Theorems 1 and 2 and the Appendix) for the BII estimator, state the asymptotic results for this estimator and sketch the proofs.

Theorem 3.1: *Assume the usual regularity conditions for the Cox proportional hazards model (Fleming & Harrington, 1991, p. 289-90). Further assume that the selection probability $\alpha(t)$ is strictly positive, $\alpha(t) > 0$ for all $t \in [0, \tau]$, and that all covariates are fixed. Then the pseudoscore* (3.1) *of the BII estimator can be decomposed into the partial likelihood score* (2.4)*, a sum of iid. zero mean terms and a remainder term as follows:*

$$\begin{aligned}\frac{1}{\sqrt{n}}\boldsymbol{U}_C(\boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\boldsymbol{U}(\boldsymbol{\beta}_0)\\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\left(1-\frac{\xi_i}{\alpha}\right)\int_0^{\tau}\{\boldsymbol{R}_i(t) - \frac{Y_i(t)}{m(t)}\psi(t)\}d\Lambda_0(t)\\ &+ o_P(1),\end{aligned}$$

where $\boldsymbol{R}_i(t) = [\boldsymbol{Z}_i - \bar{\boldsymbol{z}}]\exp\{\boldsymbol{\beta}_0'\boldsymbol{Z}_i\}Y_i(t),\ m(t) = \mathsf{E}(1-\delta_i)Y_i(t),\ \psi(t) = \mathsf{E}(1-\delta_i)\boldsymbol{R}_i(t),\ \Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$ and $\bar{\boldsymbol{z}}$ is the uniform probability limit of $\mathsf{E}(\boldsymbol{\beta}, t)$ (see Fleming & Harrington, 1991).

Proof: The proof is an adaptation of the proof given by (Kulich & Lin, 2004, Appendix A3) to the BII estimator. However, rather than a complete and rigorous proof we give a commented sketch of the proof.

The pseudoscore (3.1) can be represented as

$$\frac{1}{\sqrt{n}}\boldsymbol{U}_{\boldsymbol{C}}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\boldsymbol{U}(\boldsymbol{\beta}_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}(\mathsf{E}(\boldsymbol{\beta}_0, t) - \mathsf{E}_C(\boldsymbol{\beta}_0, t))dN_i(t).$$

The counting process $N_i(t)$ can be decomposed to a martingale $M_i(t)$ and a compensator $\int_0^t Y_i(s)\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}_i)d\Lambda(s)$, which splits the above integral into two components. The first integral with respect to $M_i(t)$ can be shown to converge to zero in probability, thus $\frac{1}{\sqrt{n}}\sum_i\int_0^{\tau}(\mathsf{E}(\boldsymbol{\beta}_0, t) - \mathsf{E}_C(\boldsymbol{\beta}_0, t))dN_i(t)$ can be approximated by

$$\frac{1}{\sqrt{n}}\int_0^{\tau}\big(\mathsf{E}(\boldsymbol{\beta}_0, t) - \mathsf{E}_C(\boldsymbol{\beta}_0, t)\big)\sum_i Y_i(t)\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}_i)d\Lambda_0(t). \tag{3.3}$$

Now we can express the integrand of (3.3) in terms of $\boldsymbol{S}^{(0)}, \boldsymbol{S}^{(1)}$ and $\boldsymbol{S}_C^{(0)}, \boldsymbol{S}_C^{(1)}$. Since $(\mathsf{E} - \mathsf{E}_C)S^{(0)} = (\boldsymbol{S}^{(1)} - \boldsymbol{S}_C^{(1)}) + \mathsf{E}_C(S_C^{(0)} - S^{(0)})$ and E_C converges uniformly to $\bar{\boldsymbol{z}}$, (3.3) becomes

$$\sqrt{n}\int_0^\tau (\boldsymbol{S}^{(1)} - \boldsymbol{S}_C^{(1)})d\Lambda_0 + \sqrt{n}\int_0^\tau (S_C^{(0)} - S^{(0)})\bar{\boldsymbol{z}}d\Lambda_0 + o_P(1). \tag{3.4}$$

Substituting the definitions of $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}_C^{(1)}$ to the first part of (3.4) we get

$$\begin{aligned}\sqrt{n}\int_0^\tau &(\boldsymbol{S}^{(1)} - \boldsymbol{S}_C^{(1)})d\Lambda_0(t) = \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1 - \frac{\xi_i}{\alpha}\right)(1-\delta_i)\int_0^\tau \boldsymbol{Z}_i \exp(\boldsymbol{\beta}_0 \boldsymbol{Z}_i)Y_i(t)d\Lambda_0(t) \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(1-\delta_i)\int_0^\tau (\hat{\alpha}^{-1} - \alpha^{-1})\boldsymbol{Z}_i \exp(\boldsymbol{\beta}_0 \boldsymbol{Z}_i)Y_i(t)d\Lambda_0(t).\end{aligned} \tag{3.5}$$

While the first part of (3.5) is already a sum of iid terms, the second term needs one more approximation. This approximation is based on an asymptotic expansion of the weights (Kulich & Lin, 2004, Appendix A1) and relies heavily on the assumption that the sampling probabilities α_i are bounded away from zero. Provided $\alpha_i > \varepsilon > 0$, the last part of (3.5) can be approximated by

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_i \xi_i(1-\delta_i)\int_0^\tau \frac{\boldsymbol{Z}_i \exp(\beta_0' \boldsymbol{Z}_i)Y_i(t)}{\alpha\, m(t)}\left\{\frac{1}{n}\sum_{j=1}^n\left(1 - \frac{\xi_j}{\alpha}\right)(1-\delta_j)Y_j(t)\right\}d\Lambda_0 \\ &\qquad = \frac{1}{\sqrt{n}}\sum_j \left(1 - \frac{\xi_j}{\alpha}\right)(1-\delta_j) \\ &\qquad\quad \times \int_0^\tau \frac{Y_j(t)}{m(t)}\left\{\frac{1}{n}\sum_i \frac{\xi_i}{\alpha}(1-\delta_i)\boldsymbol{Z}_i \exp(\beta_0'\boldsymbol{Z}_i)Y_i(t)\right\}d\Lambda_0,\end{aligned}$$

where $\left\{\frac{1}{n}\sum_i \frac{\xi_i}{\alpha}(1-\delta_i)\boldsymbol{Z}_i \exp(\beta_0'\boldsymbol{Z}_i)Y_i(t)\right\} \to \psi(t)$ uniformly in t. Similar operations can be performed on the latter term in (3.4) with $S^{(0)}$ and the proof can be completed by combining these two results and Proposition A1 in the Appendix in Kulich & Lin (2004). ■

Theorem 3.1 is a key result for the asymptotical properties of the pseudoscore BII estimator $\hat{\boldsymbol{\beta}}_B$. Using Theorem 3.1 one can directly prove consistency of $\hat{\boldsymbol{\beta}}_B$ and the asymptotic normality of U_C, which in turn implies the asymptotic normality of $\hat{\boldsymbol{\beta}}_B$ itself by Taylor expansion. We shall only state the main result in Theorem 3.2 without proof, which can be found in Kulich & Lin (2004), Appendix A4, and uses Theorem 3.1 and standard approximation techniques.

Theorem 3.2: *Under conditions assumed in Theorem 3.1,*

$$n^{-1/2}\boldsymbol{U_C}(\boldsymbol{\beta}_0) \xrightarrow{D} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I} + \boldsymbol{\Sigma}_C) \text{ and}$$
$$\sqrt{n}(\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0) \xrightarrow{D} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}^{-1} + \boldsymbol{I}^{-1}\boldsymbol{\Sigma}_C\boldsymbol{I}^{-1}),$$

where $\boldsymbol{I}$ is the limiting partial likelihood information matrix (see Fleming & Harrington, 1991),

$$\boldsymbol{\Sigma}_C = \frac{1-\alpha}{\alpha} E\left[(1-\delta_i)\int_0^\tau \{\boldsymbol{R}_i(t) - \frac{Y_i(t)}{m(t)}\psi(t)\} d\Lambda_0(t)\right]^{\otimes 2}$$

and $\boldsymbol{x}^{\otimes 2} = \boldsymbol{x}\boldsymbol{x}'$ for any vector $\boldsymbol{x}$.

The asymptotic variance of $\hat{\boldsymbol{\beta}}_B$ can be expressed as the variance of the full data partial likelihood estimator plus an extra component expressing the variability induced by the case-cohort sampling. The proof is based on Theorem 3.1 that requires $\alpha(t) > 0$ for all $t \in [0, \tau]$. In practice if α is close to zero, the case-cohort estimators based on the inverse probability weighting principle do not follow their theoretical asymptotic distributions. We will illustrate this phenomenon on a small simulation study later in this chapter.

Chen & Lo (1999) give a sketch of an alternative proof of Theorem 3.1 adapted to their modified estimators. A robust approach to the variance estimation is presented in Barlow (1994) and yet another approach to estimation of the variance of $\hat{\boldsymbol{\beta}}$, which utilizes bootstrapping, can be found in Wacholder et al. (1989).

3.3 Other regression models under the case-cohort design

All the main ideas of the case-cohort design illustrated on the proportional hazards model apply directly to the whole class of nonlinear transformation models. The subcohort and cases can be selected in the same ways as described earlier. Most authors adapt estimating equations by introducing the principle of inverse probability weighting. Therefore case-cohort estimators used in transformation models generally have similar properties as those used in the proportional hazards model.

As introduced earlier in (2.10) in Section 2.4, transformation models assume that the survival time T follows the equation $H(T) = -\boldsymbol{\beta}'\boldsymbol{Z} + \varepsilon$, where H is an unknown monotone transformation function such that $H(0) = -\infty$, $\boldsymbol{\beta}$ is a p-dimensional regression parameter, $\boldsymbol{Z}$ is a $p \times 1$ vector of covariates and ε is a random error variable with a known distribution, independent of $\boldsymbol{Z}$. Specific models of this class are obtained by choosing a particular distribution of ε. Both the proportional hazards and the proportional odds models are special cases of model (2.10). The former is obtained when ε follows the extreme-value distribution, the latter arises from the standard logistic distribution.

Although the basic ideas are similar, there is one principal difference between the proportional hazards model and any other model from the class of nonlinear transformation models. The baseline hazard function $\lambda_0(t)$, which plays the role of the nonparametric (functional) part of the model, disappears from the estimating equations with full data and also under the case-cohort design. On the contrary, the functional parameter remains in some form in all the other nonlinear transformation models and has to be estimated. With most methods, this function only needs to be estimated at failure times, which are completely observed under the case-cohort design.

Let us now take a closer look at general parameter estimation methods in nonlinear transformation models under the case-cohort design.

The procedure by Lu and Tsiatis

This procedure is a direct generalisation of Chen et al. (2002). Recall that Chen et al. were motivated by a martingale representation of the model and suggested estimating equations

$$\sum_{i=1}^{n} \int_0^\infty \boldsymbol{Z}_i [dN_i(t) - Y_i(t) d\Lambda_\varepsilon\{\boldsymbol{\beta}'\boldsymbol{Z}_i + H(t)\}] = 0,$$

$$\sum_{i=1}^{n} [dN_i(t) - Y_i(t) d\Lambda_\varepsilon\{\boldsymbol{\beta}'\boldsymbol{Z}_i + H(t)\}] = 0 \quad (t \geq 0).$$

Lu & Tsiatis (2006) proposed a modification of these estimating equations to case-cohort data by weighting with inverse sampling probabilities. The modified estimating equations have the form

$$\sum_{i=1}^{n} \int_0^\infty \boldsymbol{Z}_i \varrho_i [dN_i(t) - Y_i(t) d\Lambda_\varepsilon\{\boldsymbol{\beta}'\boldsymbol{Z}_i + H(t)\}] = 0, \tag{3.6a}$$

$$\sum_{i=1}^{n} \varrho_i [dN_i(t) - Y_i(t) d\Lambda_\varepsilon\{\boldsymbol{\beta}'\boldsymbol{Z}_i + H(t)\}] = 0 \quad (t \geq 0), \tag{3.6b}$$

where the weight ϱ_i is the inverse sampling probability for each individual in the full cohort. Specifically $\varrho_i = \delta_i + (1 - \delta_i)\xi_i/\alpha$, where ξ_i is the indicator variable for being sampled into the subcohort and $\alpha = \mathrm{P}(\xi_i = 1)$ is the probability of being sampled into the subcohort. According to the terminology introduced in Section 3.2 this estimator belongs to the class of D-estimators with constant weights. Weights chosen in this way fulfill the basic requirement of eliminating all unobserved data, while the weight of 1 is given to all cases during the whole study period regardless of their subcohort status. The Chen's estimator specified for full data as a solution to (2.12a) and (2.12b) is simply obtained by setting $\alpha = 1$, $i = 1, \dots, n$, since then $\xi_i = 1$ and $\varrho_i = 1$.

The computing algorithm mimics the procedure of Chen et al. (2002). Since all failures are observed with probability one, we can keep the notation $0 < t_1 < \cdots < t_K < \infty$ for the K observed distinct failure times from the full cohort. The algorithm again switches between estimation of H and $\boldsymbol{\beta}$ and can be summarized as follows:

Step 1: Fix $\boldsymbol{\beta}$ at an initial value $\boldsymbol{\beta}^{(0)}$.

Step 2: Compute $H^{(0)}(t_1), \dots, H^{(0)}(t_K)$ by solving the equations

$$\sum_{i=1}^{n} \varrho_i Y_i(t_k)\Lambda_\varepsilon\{H(t_k) + \boldsymbol{\beta}'\boldsymbol{Z}_i\} = 1 + \sum_{i=1}^{n} \varrho_i Y_i(t_k)\Lambda_\varepsilon\{H(t_k-) + \boldsymbol{\beta}'\boldsymbol{Z}_i\}$$

for $H^{(0)}(t_k)$, $k = 1, \dots, K$, with $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ and $H^{(0)}(t_1-) = -\infty$.

Step 3: Update the estimate of $\boldsymbol{\beta}$ by solving equation (3.6a) for $\boldsymbol{\beta}$ with H fixed at the $H^{(0)}$ obtained in Step 2.

Step 4: Set $\boldsymbol{\beta}^{(0)}$ to the new value obtained in Step 3 and repeat steps 2 and 3 until convergence is reached.

Except for certain special cases, in particular the proportional hazards model, solving the nonlinear equations in steps 2 and 3 again requires an iterative procedure. The modified Newton-Raphson algorithm with step halving performs generally quite well. Similarly to the algorithm for complete data, simplifying the above equations by using first order differences rather than derivatives increases the inaccuracy in a substantial way and should not be used for computation (Klášterecký & Kulich, 2006).

Asymptotic properties

Regularity conditions needed to prove the existence, consistency and asymptotical normality of the estimator are taken from Chen et al. (2002). In addition to conditions necessary for the martingale central limit theorem (Fleming & Harrington, 1991) they mainly involve positivity and some smoothness assumptions on the hazard rate λ_ε and the transformation function H_0, the usual restriction to a finite time interval $(0, \tau]$ and a requirement on covariates to be bounded in probability.

Although it is not stated explicitly anywhere in Lu & Tsiatis (2006), the condition of positive sampling probability $\alpha > 0$ is a necessary condition. We shall only state the main result of Lu & Tsiatis without proof. A sketch of the proof can be found in Lu & Tsiatis (2006).

Theorem 3.3: *Under suitable regularity conditions the estimator $\hat{\boldsymbol{\beta}}$ defined as a solution of* (3.6a) *and* (3.6b) *exists and is consistent. Moreover, $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed with*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathrm{N}(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^{-1'}),$$

where $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ can be consistently estimated by

$$\begin{aligned}
\hat{\boldsymbol{A}} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \varrho_i(\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t))\boldsymbol{Z}_i' \frac{\partial\lambda_\varepsilon(\hat{H}(t) + \hat{\boldsymbol{\beta}}'\boldsymbol{Z}_i)}{\partial t} Y_i(t)d\hat{H}(t),\\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{i=1}^{n}\varrho_i^2\left[\int_0^\tau (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t))d\hat{M}_i(t)\right]^{\otimes 2}\\
&\quad - \frac{1-\alpha}{\alpha}\left[\frac{1}{n}\sum_{i-1}^{n}\delta_i\int_0^\tau (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t))d\hat{M}_i(t)\right]^{\otimes 2},
\end{aligned}$$

where $\hat{M}_i(t) = N_i(t) - \int_0^\tau Y_i(s)d\Lambda_\varepsilon(\hat{H}(s) + \hat{\boldsymbol{\beta}}'\boldsymbol{Z}_i)$ and $\bar{\boldsymbol{Z}}(t)$ is a kind of weighted average of the covariates of individuals being at risk at time t, see Lu & Tsiatis (2006).

Theorem 3.3 theoretically justifies consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$ and allows constructing asymptotic confidence intervals for $\boldsymbol{\beta}$ and testing hypotheses about $\boldsymbol{\beta}$. Lu & Tsiatis (2006) investigate $\hat{\boldsymbol{\beta}}$ and $\hat{H}$ in simulation studies and report good properties of $\hat{\boldsymbol{\beta}}$ – unbiasedness, good coverage of confidence intervals and some efficiency gain over the estimator by Kong et al. (2004). The behaviour of the estimator will be explored in more detail later in this chapter.

Chen's weighted semiparametric likelihood

Nearly all of the estimation methods developed with full data were modified or generalized to the case-cohort problem with one interesting exception. There is a very limited number of papers devoted specifically to the proportional odds model for case-cohort studies. The authors either cover this model by deriving estimating equations for the whole class of nonlinear transformation models or restrict their attention to the proportional hazards model only.

The proportional odds model plays quite a prominent role in the work of Chen (2001). Chen introduces the concept of weighted semiparametric likelihood for the proportional odds model, a modified version of the likelihood (2.8) derived by Murphy et al. (1997). The controls' contributions to the likelihood are weighted by the estimated inverse sampling probabilities using $\varrho_i = \delta_i + (1-\delta_i)\xi_i/\hat{\alpha}$, yielding the weighted semiparametric likelihood function

$$\prod_{i=1}^{n}\left\{\left[\frac{1}{1+H(t)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)}\right]^{\xi_i\frac{1-\delta_i}{\hat{\alpha}}}\times\left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)dH(t)}{(1+H(t)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i))^2}\right]^{\delta_i}\right\}, \qquad (3.6)$$

where $H(t)$ is the baseline odds of failure at t, $\hat{\alpha} = n_0/(n-n_1)$ is the weight for subcohort controls and n_1 and n_0 are the number of all cases and the number of sampled controls, respectively. Comparing (3.6) to the original likelihood function (2.8) on page 9 reveals that the structure of both formulas is identical. Other estimators can again be obtained using different sampling probability estimators as weight functions. However, according to Chen (2001), using the overall proportion of controls $\hat{\alpha} = n_0/n$ as the most natural estimator is less efficient.

Chen (2001) showed the existence, consistency and asymptotic normality of the resulting estimator by following the proofs of Murphy et al. (1997) and making changes when necessary. In the latter part of the paper the results are subsequently extended to estimation of survival probabilities. The method of weighted semiparametric likelihood and its connection to the approach of Chen & Lo (1999) are then studied for the whole class of nonlinear transformation models and for the Cox model in particular. One of the most important assumptions for establishing the asymptotic properties of any estimator based on (3.6) or a similar weighted likelihood function is naturally again a positive sampling probability $\alpha > 0$.

Kong's method

The Kong's estimator, that was used as a benchmark in simulation studies by Lu & Tsiatis, is an extension of the approach of Fine et al. (1998), see p. 11. The estimator is obtained by introducing inverse sampling probability weights into the estimating equations. The original idea of estimating the survival function for the censoring variable remains unchanged but the survival function is now estimated from the case-cohort data only. The authors give methods and formulas for making inference about regression parameters, for estimating the survival function and also for its confidence bands. Since independence of the censoring variable on covariates remains to be a necessary condition, the Kong's approach is less attractive compared to e.g. the algorithm of Lu & Tsiatis (2006).

3.4 Performance of case-cohort estimators

Asymptotic results for currently used case-cohort estimators assume that the sampling probability α is positive. However, this asymptotic theory is also used in situations where few controls are sampled for the analysis. Commonly an equal number of controls and cases are sampled from a huge cohort, see e.g. a recent paper by Vogel et al. (2004) concerning risk of lung cancer where there were 265 cases and 270 subcohort controls sampled for the analysis out of a cohort consisting of 55 000 members. The sampling proportion of the controls was only around 0.005. We have reviewed this analysis and found the resulting case-cohort parameter estimates to be biased with low confidence interval coverage.

This Section is an extension of the study presented by Klášterecký & Kulich (2006). We shall examine the behaviour of two estimators used in case-cohort studies: the classical Prentice estimator for the proportional hazards model and the estimator proposed by Lu & Tsiatis (2006) for the proportional odds model. Both estimators are examined under various settings and with different values of α. The estimator by Lu and Tsiatis has been chosen because it is very general and covers the whole class of nonlinear transformation models. Similar results can however be expected with other estimators based on the usual inverse sampling probability weighting principle.

Simulation settings

In every set of simulations we generated 1000 full cohorts with complete data. Failure times were generated from the proportional hazards model for the Prentice estimator and from the proportional odds model for the estimator by Lu and Tsiatis. Censoring times were independent of the covariates and uniformly distributed over the interval $(0, c)$ where c was chosen so that we observed approximately 100 cases in each sample. The subcohort was selected by independent Bernoulli sampling so that the expected number of subcohort controls was equal to 100, i.e. to the expected number of failures.

We considered cohorts consisting of 10 000, 50 000, 100 000 and 300 000 subjects, each time with 100 cases giving a failure rate of $0,01$, $0,002$,

$0,001$ and $0,0003$, respectively. Keeping around 100 subcohort controls, we achieved subcohort sampling probabilities α to be also $0,01$, $0,002$, $0,001$ and $0,0003$, respectively. For each α we simulated a scenario with two independent covariates (i) and a scenario with three correlated covariates (ii).

The simpler scenario (i) is very frequently reported in the literature for illustrating theoretical results and was also used by Lu and Tsiatis. Two independent covariates $Z_1 \sim \mathrm{U}(0,1)$ and $Z_2 \sim \mathrm{Alt}(0.35)$ were generated, the true regression parameters were set to $\beta_1 = 1$, $\beta_2 = -1$.

The latter scenario (ii) represents one of other practically relevant settings. We considered three mutually correlated covariates: a dichotomous covariate $Z_1 \sim \mathrm{Alt}(0.35)$ and two continuous covariates. The conditional distribution of Z_2 given Z_1 was normal with mean $-0.2Z_1$ and variance $(0.5{+}0.2Z_1)^2$. The conditional distribution of Z_3 given Z_1 and Z_2 was normal with mean $-0.3Z_1 + 0.3Z_2$ and variance $(0.1 + 0.1Z_1 + 0.1|Z_2|)^2$. The conditional normal distributions for Z_2 and Z_3 were truncated 3 standard deviations away from the mean, because the asymptotic theory assumes bounded covariates. The true parameter values were set to $\beta_1 = 2.3$, $\beta_2 = 0.7$, and $\beta_3 = 2.9$.

Results

For the Prentice estimator, the results summarised in Table 3.1 confirm that the estimator performs well in scenario (i) with independent covariates and its properties are very little influenced by the subcohort sampling probability. Even with a very low value of the sampling probability $\alpha = 0,0003$, parameter estimates are only slightly biased, standard errors are well estimated and the confidence interval coverage is good.

For correlated covariates and larger parameter values, however, the performance of the Prentice's estimator is much worse. Already for $\alpha = 0.01$ the estimator suffers from substantial bias, underestimated standard errors and poor confidence interval coverage. The results get worse when we further decrease the subcohort sampling probability α.

A similar pattern can be seen for the estimator by Lu & Tsiatis. The results summarised in Table 3.2 confirm that the estimator performs well in scenario (i) with independent covariates. In all simulations the parameter estimates have only a slight bias and good confidence interval coverage. The influence of the subcohort sampling probability is more apparent than in the proportional hazards model, but still negligible.

Results for dependent covariates and larger parameter effects are rather unsatisfactory. The performance of the estimator was much worse with a clear bias, underestimated standard errors and poor confidence interval coverage. As for the proportional hazards model, the results are bad even for $\alpha = 0.01$ and get worse when the sampling rate decreases.

Table 3.1: Simulation summary for the Prentice's estimator.

(i) Independent covariates, $\boldsymbol{\beta}_1 = 1,\ \boldsymbol{\beta}_2 = -1$

α	Par.	Bias	Mean Estim.	Empirical std. err.	Mean est. std. err.	95% CI coverage
$\alpha = 0.01$, Coh. 10 000	β_1	-0.031	1.031	0.599	0.578	0.940
	β_2	0.020	-1.020	0.385	0.379	0.946
$\alpha = 0.02$, Coh. 50 000	β_1	-0.014	1.014	0.581	0.577	0.957
	β_2	0.035	-1.035	0.394	0.379	0.947
$\alpha = 0.001$, Coh. 100 000	β_1	-0.016	1.016	0.607	0.580	0.944
	β_2	0.026	-1.026	0.377	0.378	0.956
$\alpha = 0.0003$, Coh. 300 000	β_1	-0.041	1.041	0.560	0.577	0.962
	β_2	0.018	-1.018	0.391	0.378	0.937

(ii) Dependent covariates, $\boldsymbol{\beta}_1 = 2.3,\ \boldsymbol{\beta}_2 = 0.7,\ \boldsymbol{\beta}_3 = 2.9$

α	Par.	Bias	Mean Estim.	Empirical std. err.	Mean est. std. err.	95% CI coverage
$\alpha = 0.01$, Coh. 10 000	β_1	-0.201	2.501	0.687	0.488	0.825
	β_2	-0.102	0.802	0.634	0.479	0.858
	β_3	-0.286	3.286	1.457	1.029	0.833
$\alpha = 0.002$, Coh. 50 000	β_1	-0.242	2.542	0.709	0.486	0.808
	β_2	-0.108	0.808	0.677	0.492	0.861
	β_3	-0.507	3.407	1.506	1.025	0.809
$\alpha = 0.001$, Coh. 100 000	β_1	-0.235	2.535	0.736	0.491	0.815
	β_2	-0.208	0.808	0.677	0.492	0.851
	β_3	-0.474	3.374	1.600	1.040	0.786
$\alpha = 0.0003$, Coh. 300 000	β_1	-0.258	2.578	0.755	0.493	0.795
	β_2	-0.123	0.823	0.696	0.491	0.844
	β_3	-0.491	3.391	1.575	1.037	0.802

Table 3.2: Simulation summary for the estimator by Lu & Tsiatis.

(i) Independent covariates, $\boldsymbol{\beta}_1 = 1,\ \boldsymbol{\beta}_2 = -1$

α	Par.	Bias	Mean Estim.	Empirical std. err.	Mean est. std. err.	95% CI coverage
$\alpha = 0.01$,	β_1	-0.018	1.018	0.599	0.588	0.950
Coh. 10 000	β_2	0.018	-1.018	0.387	0.380	0.949
$\alpha = 0.02$,	β_1	-0.042	1.042	0.616	0.588	0.937
Coh. 50 000	β_2	0.025	-1.025	0.376	0.380	0.962
$\alpha = 0.001$,	β_1	-0.036	1.036	0.609	0.586	0.954
Coh. 100 000	β_2	0.015	-1.015	0.379	0.380	0.950
$\alpha = 0.0003$,	β_1	-0.048	1.048	0.660	0.588	0.951
Coh. 300 000	β_2	0.063	-1.063	0.388	0.382	0.955

(ii) Dependent covariates, $\boldsymbol{\beta}_1 = 2.3,\ \boldsymbol{\beta}_2 = 0.7,\ \boldsymbol{\beta}_3 = 2.9$

α	Par.	Bias	Mean Estim.	Empirical std. err.	Mean est. std. err.	95% CI coverage
$\alpha = 0.01$, Coh. 10 000	β_1	-0.378	2.678	0.760	0.525	0.772
	β_2	-0.140	0.840	0.688	0.520	0.858
	β_3	-0.676	3.576	1.611	1.100	0.759
$\alpha = 0.002$, Coh. 50 000	β_1	-0.525	2.825	0.882	0.555	0.715
	β_2	-0.163	0.863	0.830	0.560	0.819
	β_3	-1.025	3.925	1.963	1.158	0.687
$\alpha = 0.001$, Coh. 100 000	β_1	-0.547	2.847	0.982	0.572	0.705
	β_2	-0.161	0.861	0.826	0.566	0.851
	β_3	-1.152	4.052	2.029	1.191	0.660
$\alpha = 0.0003$, Coh. 300 000	β_1	-0.671	2.971	1.107	0.597	0.683
	β_2	-0.244	0.944	0.939	0.592	0.800
	β_3	-1.380	4.280	2.295	1.241	0.650

For illustration we also show histograms of the simulated estimates for the proportional odds model, $\alpha = 0.01$, on Figures 3.1 and 3.2. Each histogram is shown with the density of the respective theoretical asymptotic normal distribution for comparison. Figure 3.1 shows the situation with independent covariates: the asymptotic normal distribution well approximates the empirical distribution of the simulated estimates. Figure 3.2 (dependent covariates) reveals noticeable skewness towards the upper tail of the distribution for all the three parameter estimates.

Figure 3.1: Relative frequency histograms of simulated case-cohort estimates of (a) β_1 and (b) β_2 with independent covariates.

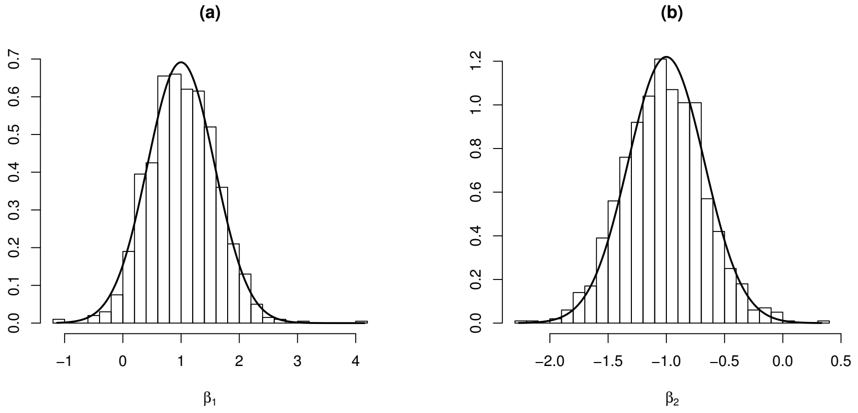
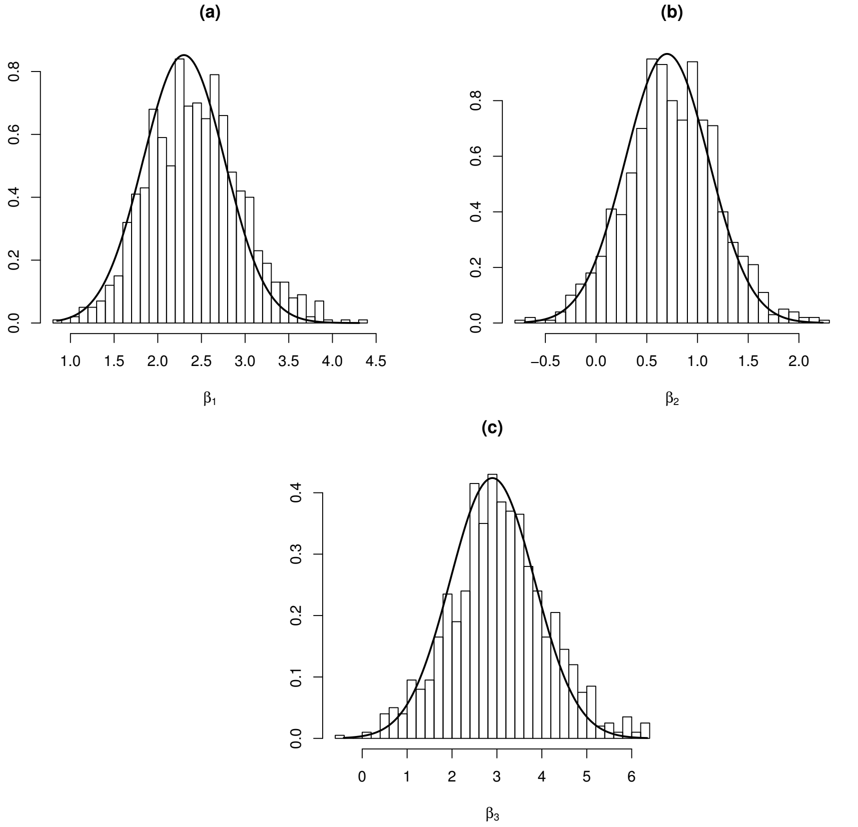


Figure 3.2: Relative frequency histograms of simulated case-cohort estimates of (a) β_1, (b) β_2, and (c) β_3 with correlated covariates.



Summary

Our simulation study shows that both estimators performed well in scenario (i) with two independent covariates. The subcohort sampling probability α only had a minor effect on the properties of both estimators. Similar results are usually also reported for other case-cohort estimators in the literature. Both estimators behaved much worse in scenario (ii) with dependent covariates. Even for the best case with $\alpha = 0.01$ the estimators were biased and had underestimated standard errors. We can also see the influence of α on the results, smaller values of α resulted in worse properties of both estimators.

Our simulation study focused only on the estimators proposed by Prentice and Lu & Tsiatis. However, any case-cohort estimator utilising the same principle of weighting by inverse sampling probabilities is likely to be affected by similar problems. We have seen that even values of α around 0.01 are small enough and theoretical results can break down here. In practice the sampling fractions can be considerably lower than 1% of the whole cohort, leading to even worse properties of known estimators. The full cohort can include hundreds of thousands of people, while the number of cases remains in the hundreds and the size of the subcohort comparable to the number of cases or somewhat larger.

The idea and principles of the case-cohort design are most useful under such rare-event scenarios and developing a good estimator is therefore very important. Constructing new estimators in the usual manner by modifying the existing weights does not solve the problem because the principle remains unchanged. In this work, we shall introduce an alternative estimation method based on logistic regression models.

3.5 Logistic models and case-control data

In Section 2.5 we have briefly introduced the logistic regression model as a tool for analysing cross-sectional data with a binary response variable. When dealing with case-cohort data, an alternative approach is using a case-control logistic regression analysis at the end of the study. Case-control data can be easily obtained from a case-cohort study by counting the numbers of cases and controls and ignoring the actual times of events and censoring.

A logistic regression model can be fitted to retrospectively collected case-control data without the need for inverse probability weighting (Prentice & Pyke, 1979). We shall exploit this property of logistic regression when developing a new estimator in Chapter 4. In this section we review the relationship between prospective and retrospective models and briefly summarise main results of Prentice & Pyke (1979). Their paper is very general but we only focus on developing the retrospective likelihood function and asymptotic properties of the odds ratio estimator.

Likelihood and score for logistic models in retrospective studies

Logistic regression models were originally developed for prospective data, that is for observations sampled from the conditional distribution of event (disease) status given the covariates. In our notation[2] this means sampling from the distribution of δ given the covariates $\boldsymbol{Z}$ – let us denote it[3] by $\mathrm{P}(\delta|\boldsymbol{Z})$. In case-control studies, however, data are sampled from the distribution of covariates $\boldsymbol{Z}$ given the event status δ (denoted by $\mathrm{P}(\boldsymbol{Z}|\delta)$).

Prentice & Pyke (1979) started from the original prospective logistic regression model (2.13) written as

$$\begin{aligned} \mathrm{P}(\delta=1|\boldsymbol{Z}) &= \frac{\exp(\alpha+\boldsymbol{\beta}'\boldsymbol{Z})}{1+\exp(\alpha+\boldsymbol{\beta}'\boldsymbol{Z})}, \\ \mathrm{P}(\delta=0|\boldsymbol{Z}) &= \frac{1}{1+\exp(\alpha+\boldsymbol{\beta}'\boldsymbol{Z})}. \end{aligned} \tag{3.7}$$

The (prospective) odds ratios based on (3.7) comparing an individual with covariates $\boldsymbol{Z}$ to an individual with a baseline or reference set of covariates $\boldsymbol{Z}_0$ equal

$$\begin{aligned} \frac{\mathrm{P}(\delta=1|\boldsymbol{Z})/\mathrm{P}(\delta=0|\boldsymbol{Z})}{\mathrm{P}(\delta=1|\boldsymbol{Z}_0)/\mathrm{P}(\delta=0|\boldsymbol{Z}_0)} &= \exp(\boldsymbol{\beta}'(\boldsymbol{Z}-\boldsymbol{Z}_0)) \\ &= \frac{\mathrm{P}(\boldsymbol{Z}|\delta=1)/\mathrm{P}(\boldsymbol{Z}_0|\delta=1)}{\mathrm{P}(\boldsymbol{Z}|\delta=0)/\mathrm{P}(\boldsymbol{Z}_0|\delta=0)}. \end{aligned} \tag{3.8}$$

The last equality is true due to the fact that $\mathrm{P}(\delta|\boldsymbol{Z}) = \mathrm{P}(\boldsymbol{Z}|\delta)\mathrm{P}(\delta)/\mathrm{P}(\boldsymbol{Z})$ and gives the first important result: the odds ratios can be estimated from retrospective data. Furthermore, using (3.8) we can write

$$\begin{aligned} &\mathrm{P}(\boldsymbol{Z}|\delta=1) = c_1\exp(\gamma(\boldsymbol{Z})+\boldsymbol{\beta}'\boldsymbol{Z}) \text{ and} \\ &\mathrm{P}(\boldsymbol{Z}|\delta=0) = c_0\exp(\gamma(\boldsymbol{Z})), \end{aligned} \tag{3.9}$$

where $\gamma(\boldsymbol{Z}) = \log[\mathrm{P}(\boldsymbol{Z}|\delta=0)/\mathrm{P}(\boldsymbol{Z}_0|\delta=0)]$ and $c_0(\gamma)$ and $c_1(\gamma,\boldsymbol{\beta})$ are normalizing factors. So, the resulting model for retrospective data is again of

[2] The censoring indicator δ becomes the response variable for the case-control analysis.

[3] During this section we shall use a unified notation by $\mathrm{P}(\cdot)$ for both, discrete and continuous distributions.

the logistic form with γ in place of the intercept term. This new model is in fact induced by the original model (2.13) which would be fitted to prospective data and both models are equivalent if the original α in (2.13) and the function γ in (3.9) are both unrestricted.

The likelihood function for the retrospective model is given by a product of corresponding parts of (3.9) over all observations. It is more convenient to work with a reparametrized problem where we put

$$q(\boldsymbol{Z}) = \exp(\gamma(\boldsymbol{Z}))[\frac{n_0}{n_0+n_1}c_0 + \frac{n_1}{n_0+n_1}c_1 \exp(\boldsymbol{\beta}'\boldsymbol{Z})]. \tag{3.10}$$

By n_0 and n_1 we denote the number of controls and cases in the sample, respectively. The function $q(\cdot)$ introduced in (3.10) can be interpreted as the marginal probability density function for $\boldsymbol{Z}$ under the case-control sampling scheme with $\mathrm{P}(\delta = i) = n_i/(n_1+n_0)$, $i = 0, 1$. To simplify the formulas let us finally introduce $\eta_i = \log(c_i n_i/(n_0+n_1))$, $i = 0, 1$. With these reparametrisations we arrive to a likelihood function L_{LR} proportional to

$$\begin{aligned} L_{\mathrm{LR}} \propto & \prod_{j=1}^{n_0} \frac{\exp(\eta_0)}{\exp(\eta_0)+\exp(\eta_1+\boldsymbol{Z}_j'\boldsymbol{\beta})} \prod_{j=1}^{n_1} \frac{\exp(\eta_1+\boldsymbol{Z}_j'\boldsymbol{\beta})}{\exp(\eta_0)+\exp(\eta_1+\boldsymbol{Z}_j'\boldsymbol{\beta})} \prod_{j=1}^{n_0+n_1} q(\boldsymbol{Z}_j) \\ & = L_1(\eta_0,\eta_1,\boldsymbol{\beta}) \times \prod_{j=1}^{n_0+n_1} q(\boldsymbol{Z}_j). \end{aligned} \tag{3.11}$$

When we treat $q(\cdot)$ as a functional nuisance parameter, the likelihood (3.11) depends on the remaining model parameters solely through L_1 and any inference concerning the regression parameters can thus be now based on L_1 only. Further calculations show that

$$\begin{aligned} \log L_1(\eta_0,\eta_1,\boldsymbol{\beta}) = & \sum_{j=1}^{n_0} [\eta_0 - \log(\exp(\eta_0)+\exp(\eta_1+\boldsymbol{Z}_j'\boldsymbol{\beta}))] \\ & + \sum_{j=1}^{n_1} [\eta_0 + \boldsymbol{Z}_j'\boldsymbol{\beta} - \log(\exp(\eta_0)+\exp(\eta_1+\boldsymbol{Z}_j'\boldsymbol{\beta}))] \\ = & \, n_1(\eta_1-\eta_0) + \sum_{j=1}^{n_1} \boldsymbol{Z}_j'\boldsymbol{\beta} \\ & - \sum_{j=1}^{n_0+n_1} \log(1+\exp(\eta_1-\eta_0+\boldsymbol{Z}_j'\boldsymbol{\beta})), \end{aligned}$$

leading to score equations

$$\mathbf{0} = \frac{\partial \log L_1}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_1} \boldsymbol{Z}_j - \exp(\eta_1 - \eta_0) \sum_{j=1}^{n_0+n_1} \frac{\boldsymbol{Z}_j \exp(\boldsymbol{Z}_j' \boldsymbol{\beta})}{1 + \exp(\eta_1 - \eta_0)\exp(\boldsymbol{Z}_j' \boldsymbol{\beta})},$$

$$0 = \frac{\partial \log L_1}{\partial \eta_i} = \pm \left(n_1 - \sum_{j=1}^{n_0+n_1} \frac{\exp(\eta_1 - \eta_0)\exp(\boldsymbol{Z}_j' \boldsymbol{\beta})}{1 + \exp(\eta_1 - \eta_0)\exp(\boldsymbol{Z}_j' \boldsymbol{\beta})} \right). \tag{3.12}$$

Solving (3.12) simultaneously provides estimates of regression parameters. Note that the main parameter of interest from the prospective model (3.7), $\boldsymbol{\beta}$, remains exactly the same parameter here. Note also that since we have case-control data, we cannot make any population-wide inference regarding the intercept term or anything related to the intercept. We cannot for example estimate the overall probability of failure from case-control data, no matter how they were collected.

Asymptotic results

The asymptotic theory for retrospective case-control parameter estimators requires some nonstandard approaches due to the nuisance functional parameter q, but the main results are quite straightforward. Denote by $\boldsymbol{\theta} = (\eta_0, \eta_1, \boldsymbol{\beta})^T$ all the parameters, let

$$G(\boldsymbol{\theta}) = \mathsf{E}\left\{ -\frac{1}{n} \frac{\partial^2 \log L_1}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right\}$$

be the expected information matrix based on L_1 and denote the normalised score by

$$S(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \frac{\partial \log L_1}{\partial \boldsymbol{\theta}}.$$

Prentice & Pyke (1979) showed that $S(\boldsymbol{\theta})$ is asymptotically normal with mean zero and variance matrix

$$\boldsymbol{\Sigma}_{cc} = \mathsf{E}\left\{ \left(\frac{\partial \log L_1}{\partial \boldsymbol{\theta}} |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \left(\frac{\partial \log L_1}{\partial \boldsymbol{\theta}} |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^T \right\}, \tag{3.13}$$

where $\boldsymbol{\theta}_0$ denotes the true parameter values in the *retrospective* model. Using Taylor expansion of the first derivative of $\log L_1$ around $\boldsymbol{\theta}_0$ then provides the asymptotic normality for the case-control regression estimators and their asymptotic variance, which equals

$$G^{-1}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}_{cc} G^{-1}(\boldsymbol{\theta}_0). \tag{3.14}$$

Prentice & Pyke (1979) also showed that the bottom-right submatrix of (3.14) related to $\boldsymbol{\beta}$ is exactly equal to the corresponding bottom-right submatrix of $G^{-1}(\boldsymbol{\theta}_0)$, which would be the asymptotic variance matrix from the prospective logistic regression model obtained by standard maximum likelihood estimation theory. This asymptotic variance matrix can be consistently estimated from retrospective data by plugging in the estimated parameters. The resulting submatrix for $\boldsymbol{\beta}$ remains correct even though the intercept term has a different meaning.

In summary, the most important message is that the same formal model can be used for analysing data stemming from both types of studies, prospective as well as retrospective. The case-control estimator of the odds ratio remains consistent and asymptotically normal, its asymptotic variance matrix is the same as if the model were applied to prospectively collected data and the estimated parameters can be plugged in to obtain a consistent estimator of the variance matrix of the odds ratios.

Chapter 4

The Combined Logistic Estimator

In Section 3.4 we have seen that the usual case-cohort estimators may encounter serious performance problems. Two estimators suggested by Prentice and by Lu and Tsiatis were used as examples, however similar problems affect all currently known estimators that are based on the inverse probability weighting principle. It is therefore desirable to develop a new estimator with better properties, especially for situations with very low event probabilities. The new estimator is presented in the current chapter. First we show how the likelihood and score functions of the proportional odds model can be rewritten into a form comparable to the likelihood and score of the logistic regression model; we also highlight similar aspects and important differences resulting from such comparison. In Section 4.2 we present the main ideas and construct the new estimator and in Section 4.3 we formulate and prove its asymptotic properties.

4.1 Background

Logistic regression models can estimate odds ratios from retrospective case-control data without weighting the observations by inverse sampling probabilities (see Section 3.5 or Prentice & Pyke (1979) for more details). We aim to avoid using inverse sampling probabilities for reasons explained in Chapter 3, so we view the logistic regression model as a natural alternative to traditional approaches.

Among all survival regression models, the proportional odds model is closest to logistic regression in many aspects, as we will show in more detail later in this section. Therefore we developed the new estimator for data following the proportional odds model and used the close relationship of proportional odds and logistic models to construct the estimator.

In situations with very low event probabilities, in which we are particularly interested, the proportional odds model closely approximates the Cox model because odds ratios closely approximate hazard ratios. In such circumstances, the proportional odds model can be used to analyse data generated by the Cox model (see the simulation study in Chapter 5).

Logistic regression and the proportional odds model

Logistic regression and the proportional odds model both estimate odds ratios (or log odds ratios). The logistic analysis does this only at one, fixed time point. The proportional odds model takes time dependence into account through the baseline log odds function $G(t)$, which replaces the intercept term from logistic regression. There is no time dependence for $\boldsymbol{\beta}$, the model assumes that the effects of model covariates do not change over time. The proportional odds model can thus be treated as a direct generalization of the logistic regression model allowing time-dependent intercepts.

Similarly, the case-cohort design can be viewed as a direct generalization of the case-control design – a case-control study can be obtained from a case-cohort study by simply ignoring or not observing the actual failure times, their ranks or other statistics. In other words, a case-control study occurs if $\delta = \mathbb{I}_{[T \leq C]}$ is treated as the response variable and controls are sampled instead of following all individuals.

This leads to the idea that the proportional odds model, applied to case-cohort data, could provide consistent parameter estimators just like logistic regression does in case-control studies. A naive approach would simply fit the proportional odds model to case-cohort data as if all individuals were observed. This is the way case-control data are analysed via logistic regression, but this logistic regression is performed at a single given time point. Viewed only at this one time point, cases and controls remain the same during the whole study. On the other hand, subjects may change their status from controls to cases when they are followed in a case-cohort study. For a more formal illustration of the differences we need to compare the odds ratios, likelihoods and score functions from both models.

Odds ratios and sampling probabilities

Denote by p the probability of developing the event of interest (or being a case), given covariates: $p = \mathrm{P}(\delta = 1|\boldsymbol{Z} = \boldsymbol{z})$. For the probability of being a control we thus have $\mathrm{P}(\delta = 0|\boldsymbol{Z} = \boldsymbol{z}) = 1 - p$. Denote by $\xi = 1$ the event that an observation is sampled for the analysis. With full data we would have $\mathrm{P}(\xi = 1) = 1$ for all cohort members independently of event status, implying $\mathrm{P}(\delta = 1|\boldsymbol{z}, \xi = 1) = \mathrm{P}(\delta = 1, \xi = 1|\boldsymbol{z})/\mathrm{P}(\xi = 1) = \mathrm{P}(\delta = 1|\boldsymbol{z}) = p$. The logistic regression model with full data has the form

$$\log \frac{\mathrm{P}(\delta = 1|\boldsymbol{z})}{1 - \mathrm{P}(\delta = 1|\boldsymbol{z})} = \log \frac{p}{1-p} = \alpha + \boldsymbol{\beta}'\boldsymbol{z}.$$

In a case-control study we usually have different values of $\mathrm{P}(\xi = 1|\delta = 1)$ and $\mathrm{P}(\xi = 1|\delta = 0)$. Even if we keep the sampling independent of covariates (e.g we do not consider any stratification) we have

$$\begin{aligned}\mathrm{P}(\delta = 1|\boldsymbol{z}, \xi = 1) &= \frac{\mathrm{P}(\delta = 1, \xi = 1|\boldsymbol{z})}{\mathrm{P}(\xi = 1|\boldsymbol{z})} \\ &= \frac{p \cdot \mathrm{P}(\xi = 1|\delta = 1, \boldsymbol{z})}{p \cdot \mathrm{P}(\xi = 1|\delta = 1, \boldsymbol{z}) + (1-p) \cdot \mathrm{P}(\xi = 1|\delta = 0, \boldsymbol{z})}\end{aligned}$$

and therefore

$$\begin{aligned}\log \frac{\mathrm{P}(\delta = 1|\boldsymbol{z}, \xi = 1)}{1 - \mathrm{P}(\delta = 1|\boldsymbol{z}, \xi = 1)} &= \log \frac{p \cdot \mathrm{P}(\xi = 1|\delta = 1, \boldsymbol{z})}{(1-p) \cdot \mathrm{P}(\xi = 1|\delta = 0, \boldsymbol{z})} \\ &= \log \frac{p}{1-p} + \log \frac{\mathrm{P}(\xi = 1|\delta = 1, \boldsymbol{z})}{\mathrm{P}(\xi = 1|\delta = 0, \boldsymbol{z})} \\ &= \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \log \frac{\mathrm{P}(\xi = 1|\delta = 1, \boldsymbol{z})}{\mathrm{P}(\xi = 1|\delta = 0, \boldsymbol{z})} \\ &= \alpha^\star + \boldsymbol{\beta}'\boldsymbol{z}.\end{aligned} \tag{4.1}$$

This is the reason why the intercept term cannot be directly estimated from case-control data even when the odds ratios remain unchanged. Consequently, no function of the intercept, such as the overall event probability $\mathrm{P}(\delta = 1|\boldsymbol{z})$, can be estimated from case-control data.

For a case-cohort study assume that there is a time $\tau > 0$, which is the end of the study, and recall that $F(t)$ denotes the cumulative distribution function of T. The event of interest $[\delta = 1]$ now depends on time and becomes $[\delta(t) = \mathbb{I}_{[T \le t]}]$ with probability $\mathrm{P}[\delta(t) = 1] = p(t) = F(t)$. Since the event of interest can develop after t and before τ for an individual, who is a control at t, being a control at t does not necessarily mean being a control later.

Suppose that the cases are sampled at their failure times and not before. Then sampling (observing an individual) is also time-dependent and $[\xi(t) = 1]$ denotes the event of being sampled at time t. The conditional probabilities and odds ratios are

$$\begin{aligned}\mathrm{P}(\delta(t)=1|\boldsymbol{z},\xi(t)=1) &= \frac{\mathrm{P}(\delta(t)=1,\xi(t)=1|\boldsymbol{z})}{\mathrm{P}(\xi(t)=1|\boldsymbol{z})} = \frac{\mathrm{P}(T\le t,\xi(t)=1|\boldsymbol{z})}{\mathrm{P}(\xi(t)=1|\boldsymbol{z})}\\ &= \frac{\mathrm{P}(\xi(t)=1|T\le t,\boldsymbol{z})\cdot F(t)}{\mathrm{P}(\xi(t)=1|T\le t,\boldsymbol{z})\cdot F(t)+\mathrm{P}(\xi(t)=1,T>t|\boldsymbol{z}))}.\end{aligned}$$

The last term in the denominator can be rewritten as

$$\begin{aligned}\mathrm{P}(\xi(t)=1,T>t|\boldsymbol{z}) &= \mathrm{P}(\xi(t)=1|t<T\le\tau|\boldsymbol{z})\cdot\mathrm{P}(t<T\le\tau)\\ &\quad+\mathrm{P}(\xi(t)=1|T>\tau|\boldsymbol{z})\cdot\mathrm{P}(T>\tau)\\ &= \mathrm{P}(\xi(t)=1|t<T\le\tau|\boldsymbol{z})\cdot[F(\tau)-F(t)]\\ &\quad+\mathrm{P}(\xi(t)=1|T>\tau|\boldsymbol{z})\cdot[1-F(\tau)],\end{aligned}$$

where the extra term $\mathrm{P}(\xi(t)=1|t<T\le\tau|\boldsymbol{z}))$ is the probability of sampling a control at t that becomes a case prior to τ. Therefore

$$\log\frac{\mathrm{P}(T\le t|\boldsymbol{z},\xi(t)=1)}{1-\mathrm{P}(T\le t|\boldsymbol{z},\xi(t)=1)} = \log\frac{F(t)\cdot\mathrm{P}(\xi(t)=1|T\le t,\boldsymbol{z})}{D(t,\tau)}, \tag{4.2}$$

where

$$\begin{aligned}D(t,\tau) &= [1-F(\tau)]\cdot\mathrm{P}(\xi(t)=1|T>\tau,\boldsymbol{z})\\ &\quad+[F(\tau)-F(t)]\cdot\mathrm{P}(\xi(t)=1|t<T\le\tau,\boldsymbol{z}).\end{aligned}$$

The problem is that (4.2) cannot be easily split into the original odds ratio plus a correction term like in (4.1). As time t approaches the end of study τ, the probability that a control sampled at t becomes a case prior to τ converges to zero and the odds ratio gets closer to that from a case-control study.

Likelihood and score in the proportional odds model

Let us now explore in detail the likelihood and score functions of the proportional odds model and compare them to the retrospective logistic likelihood and score from Section 3.5. Recall that under the proportional odds model we can observe the censored failure time Y_i (the response variable), censoring indicator δ_i and covariates $\boldsymbol{Z}_i, i=1,\dots n$. Murphy et al. (1997) derived that the likelihood is proportional to L_{PO}, which can be written as

$$\begin{aligned} L_{\mathrm{PO}}(H,\boldsymbol{\beta}) &= \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})h(y_i)}{(1+H(y_i)\exp(\boldsymbol{Z}_i'\boldsymbol{\beta}))^2}\right]^{\delta_i}\left[\frac{1}{1+H(y_i)\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})}\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})h(y_i)}{1+H(y_i)\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})}\right]^{\delta_i}\left[\frac{1}{1+H(y_i)\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})}\right], \end{aligned} \tag{4.3}$$

where $H(t) = \exp\{G(t)\}$ is the baseline odds of failure by time t and $h(t)$ is the first derivative of $H(t)$.

In order to compare the score equations for $\boldsymbol{\beta}$ based on (4.3) to those from logistic regression, we should rewrite (4.3) in some form that reflects its evolution over time. Suppress for a moment its dependence on H and $\boldsymbol{\beta}$ and denote it just $L_{\mathrm{PO}}(t)$. Let there be K observed failures indexed by $f_1, \ldots, f_K$ and ordered so that $t_{f_1} < t_{f_2} <, \ldots, < t_{f_K}$. For all $t < t_{f_1}$ we have $L_{\mathrm{PO}}(t) = 1$ since $H(t) = 0$ for $t < t_{f_1}$ and further

$$L_{\mathrm{PO}}(t_{f_1}) = 1 \times \frac{h(t_{f_1})\exp(\boldsymbol{Z}'_{f_1}\boldsymbol{\beta})}{1+H(t_{f_1})\exp(\boldsymbol{Z}'_{f_1}\boldsymbol{\beta})} \times \prod_{i=f_1}^{n}\frac{1}{1+H(t_{f_1})\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})},$$

$$L_{\mathrm{PO}}(t) \quad = L(t_{f_1}) \text{ for } t \in (t_{f_1}, t_{f_2}),$$

$$\begin{aligned} L_{\mathrm{PO}}(t_{f_2}) = 1 &\times \frac{h(t_{f_1})\exp(\boldsymbol{Z}'_{f_1}\boldsymbol{\beta})}{1+H(t_{f_1})\exp(\boldsymbol{Z}'_{f_1}\boldsymbol{\beta})} \times \prod_{i=f_1}^{f_2-1}\frac{1}{1+H(t_{f_1})\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})} \\ &\times \frac{h(t_{f_2})\exp(\boldsymbol{Z}'_{f_2}\boldsymbol{\beta})}{1+H(t_{f_2})\exp(\boldsymbol{Z}'_{f_2}\boldsymbol{\beta})} \times \prod_{i=f_2}^{n}\frac{1}{1+H(t_{f_2})\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})}, \end{aligned}$$

$$L_{\mathrm{PO}}(t) \quad = L(t_{f_2}) \text{ for } t \in (t_{f_2}, t_{f_3}) \text{ and so on.}$$

If we continue this way until the last observation is included, we obtain

$$L_{\mathrm{PO}}(t_K) = \prod_{k=1}^{K}\left[\frac{h(t_{f_k})\exp(\boldsymbol{Z}'_{f_k}\boldsymbol{\beta})}{1+H(t_{f_k})\exp(\boldsymbol{Z}'_{f_k}\boldsymbol{\beta})} \times \prod_{i=f_k}^{f_{k+1}-1}\frac{1}{1+H(t_{f_k})\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})}\right], \tag{4.4}$$

where we set $f_{K+1} = n+1$ for brevity. Finally, taking the logarithm and first derivative of (4.4) yields the score equation for estimating $\boldsymbol{\beta}$

$$\begin{aligned}\mathbf{0} &= \frac{\partial \log(L_{\mathrm{PO}}(t_K))}{\partial \boldsymbol{\beta}} \\ &= \sum_{k=1}^{K} \boldsymbol{Z}_{f_k} - \sum_{k=1}^{K} \left[\frac{H(t_{f_k})\boldsymbol{Z}_{f_k} \exp(\boldsymbol{Z}'_{f_k}\boldsymbol{\beta})}{1 + H(t_{f_k}) \exp(\boldsymbol{Z}'_{f_k}\boldsymbol{\beta})} + \sum_{i=f_k}^{f_{k+1}-1} \frac{H(t_{f_k})\boldsymbol{Z}_i \exp(\boldsymbol{Z}'_i\boldsymbol{\beta})}{1 + H(t_{f_k}) \exp(\boldsymbol{Z}'_i\boldsymbol{\beta})} \right] \\ &= \sum_{k=1}^{K} \frac{\boldsymbol{Z}_{f_k}}{1 + H(t_{f_k}) \exp(\boldsymbol{Z}'_{f_k}\boldsymbol{\beta})} - \sum_{k=1}^{K} \sum_{i=f_k}^{f_{k+1}-1} \frac{H(t_{f_k})\boldsymbol{Z}_i \exp(\boldsymbol{Z}'_i\boldsymbol{\beta})}{1 + H(t_{f_k}) \exp(\boldsymbol{Z}'_i\boldsymbol{\beta})}. \end{aligned} \tag{4.5}$$

Comparing (4.5) to the logistic regression score (3.12) shows that both functions look very similar. There is a correction term for the cases in (4.5) and the time-varying term $H(t_{f_k})$ is replaced by a constant term $\exp(\eta_1 - \eta_0)$ in (3.12). In fact, $H(t_{f_k})$ represents a functional parameter that captures the time-dependent part of the model. The risk set changes between two successive events and we would estimate incorrect parameters by fitting a proportional odds model directly to case-cohort data. On the other hand, logistic regression can only analyse data collected at one given time point, does not need to handle time dependence and can be applied to case-control data for estimating the odds ratios.

4.2 Construction of the estimator

Motivated by the fact that logistic regression can be used with case-control data without inverse probability weighting, the main idea of our approach is to estimate the odds ratios repeatedly by applying the logistic regression model. We obtain a sequence of estimators computed at different time points and combine them into a new estimator. The estimator will be introduced in this section with emphasis placed on main thoughts and steps in the development of the estimator. The whole section is therefore rather informal, providing a step by step explanation of the ideas and their generalizations. Rigorous proofs and detailed calculations are provided later in Section 4.3, a simulation study is presented in Chapter 5 and a discussion on some open problems follows in Chapter 6.

Combining estimators computed at fixed times

Let the data satisfy the proportional odds model

$$-\mathrm{logit}(S(t|\boldsymbol{Z})) = \alpha_0(t) + \boldsymbol{\beta}'_0\boldsymbol{Z}. \tag{4.6}$$

In Section 2.3, the same model was introduced in equation (2.6) with $G(t)$ in place of $\alpha_0(t)$. Assume that there is a time point τ, the endpoint of the study, such that all individuals that have not failed until τ are censored at τ and also assume that no other censoring occurs. A logistic regression model can be used to estimate $\boldsymbol{\beta}_0$ at any fixed time point t, $0 < t \leq \tau$. Choose K fixed time points $t_1, \dots t_K$ such that $0 < t_1 < t_2 \cdots < t_K \leq \tau$ and perform the case-control analysis at these time points $t_1, \dots t_K$. We obtain a sequence of case-control estimators $\hat{\alpha}(t_1), \hat{\boldsymbol{\beta}}(t_1), \dots, \hat{\alpha}(t_K), \hat{\boldsymbol{\beta}}(t_K)$, where each $\hat{\boldsymbol{\beta}}(t_k)$ is a consistent and asymptotically normal estimator of the odds ratios from the original proportional odds model. These estimators are computed from data available at each time of analysis. As we move t towards τ, we observe new cases and some controls develop the event of interest and become cases.

Although the latest possible analysis conducted at τ is the best one measured by the amount of information available, the earlier analyses contain additional information on the ordering of individual failures. Such information cannot be captured in any single cross-sectional analysis, not even in the latest one performed at τ. To exploit the time information contained in the sequence of estimators, these individual estimators will be combined into a single estimator of $\boldsymbol{\beta}_0$ by computing a weighted average. Let us call the new estimator *the combined logistic estimator (CLE).*

Choosing the appropriate combination

Denote by $\boldsymbol{I}$ the identity matrix and by $\boldsymbol{W}(t_1, \dots, t_K)$ a set of $p \times p$ matrices of known constants $\{\boldsymbol{W}_1, \dots, \boldsymbol{W}_K\}$ such that $\sum_{k=1}^{K} \boldsymbol{W}_k = \boldsymbol{I}_{p\times p}$. For any set $\boldsymbol{W}(t_1, \dots, t_K)$ define the combined logistic estimator as

$$\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}(t_1,\dots,t_K)} = \sum_{k=1}^{K} \boldsymbol{W}_k \, \hat{\boldsymbol{\beta}}(t_k). \tag{4.7}$$

We would like to choose $\boldsymbol{W}^{\text{opt}}(t_1, \dots, t_K)$ so that the asymptotic variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{W}^{\text{opt}}(t_1,\dots,t_K)}$ of $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}^{\text{opt}}(t_1,\dots,t_K)}$ satisfies $\{\boldsymbol{\Sigma}_{\boldsymbol{W}(t_1,\dots,t_K)} - \boldsymbol{\Sigma}_{\boldsymbol{W}^{\text{opt}}(t_1,\dots,t_K)}\} \geq \mathbf{0}$ for any choice of $\boldsymbol{W}(t_1, \dots, t_K)$. In theory, such a set of weighting matrices can be identified. However, because the optimal weighting matrices need to be estimated from the data, such estimators do not behave well even for $K = 2$ (Kulich & Lin (2004)). In our situation K is typically much larger.

Thus, we consider only diagonal matrices

$$\boldsymbol{W}_{k_{(p\times p)}} = \text{diag}\{\boldsymbol{w}_k\} = \text{diag}\{(w_{1k}, \dots, w_{pk})'\}, \ k = 1, \dots, K,$$

where $0 \leq w_{jk} \leq 1$ for all j and k. Using diagonal matrices leads to combining individual components of the parameter vector $\boldsymbol{\beta}$ separately and discards $Kp(p-1)$ variance-covariance parameters, but we are still working with highly correlated consecutive logistic regression estimators. As a consequence the influence of one estimator is often eliminated by another one (they both have weights of similar magnitudes with opposite signs). Therefore we need to restrict the weights further and work only with convex linear combinations of the individual components. The combined estimator $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}(t_1,\dots,t_K)} = (\tilde{\beta}_{1,\boldsymbol{W}(t_1,\dots,t_K)}, \dots, \tilde{\beta}_{p,\boldsymbol{W}(t_1,\dots,t_K)})'$ can be written component-wise as

$$\tilde{\beta}_{j,\boldsymbol{W}(t_1,\dots,t_K)} = \sum_{k=1}^{K} w_{jk}\hat{\beta}_j(t_k), \qquad \sum_{k=1}^{K} w_{jk} = 1, \qquad 0 \leq w_{jk} \leq 1. \tag{4.8}$$

It can be shown (see Section 4.3), that any estimator $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}(t_1,\dots,t_K)}$ belonging to the class defined in (4.7) or (4.8) is consistent and asymptotically normally distributed,

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}(t_1,\dots,t_K)} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{W}(t_1,\dots,t_K)}), \tag{4.9}$$

where the diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{W}(t_1,\dots,t_K)}$ are given by

$$\mathsf{V}(\tilde{\beta}_j) = \sum_{k=1}^{K} w_{jk}^2 \,\mathsf{V}(\hat{\beta}_j(t_k)) + 2\sum_{k=1}^{K}\sum_{l=k+1}^{K} w_{jk}w_{jl}\,\mathsf{C}(\hat{\beta}_j(t_k), \hat{\beta}_j(t_l)) \tag{4.10}$$

and the off-diagonal elements by

$$\begin{aligned}\mathsf{C}(\tilde{\beta}_j, \tilde{\beta}_{j'}) =&\ \mathsf{C}\Big(\sum_{k=1}^{K} w_{jk}\hat{\beta}_j(t_k), \sum_{l=1}^{K} w_{j'l}\hat{\beta}_{j'}(t_l)\Big)\\ =& \sum_{k=1}^{K} w_{jk}w_{j'k}\,\mathsf{C}(\hat{\beta}_j(t_k), \hat{\beta}_{j'}(t_k)) + \sum_{k=1}^{K}\sum_{l=1,\, l\neq k}^{K} w_{jk}w_{j'l}\,\mathsf{C}(\hat{\beta}_j(t_k), \hat{\beta}_{j'}(t_l)).\end{aligned} \tag{4.11}$$

By $\mathsf{V}(\cdot)$ and $\mathsf{C}(\cdot,\cdot)$ we denote asymptotic variances and covariances of the arguments.

Intuitively, consistency is clear since each $\hat{\boldsymbol{\beta}}(t_k)$ is a consistent estimator of $\boldsymbol{\beta}_0$ and $\sum_{k=1}^{K} \boldsymbol{W}_k = \boldsymbol{I}_{p\times p}$. The asymptotic normality follows from the joint asymptotic normality of the vector of estimators $(\hat{\boldsymbol{\beta}}(t_1), \dots, \hat{\boldsymbol{\beta}}(t_K))'$, which is also proved in Section 4.3. More detailed expressions for the individual variances and covariances can be obtained through logistic regression score functions and are given in Theorem 4.4 and Corollary 4.5.

The *optimal combined logistic estimator* $\tilde{\boldsymbol{\beta}}^{\mathrm{opt}}_{\boldsymbol{W}(t_1,\dots,t_K)}$ is defined through weights that minimize (4.10), the asymptotic variance of each component of $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}(t_1,\dots,t_K)}$. The constrained minimization of variances (4.10) for each $j = 1, \dots, p$ with respect to $w_{j1}, \dots, w_{jK}$ subject to $\sum_{k=1}^{K} w_{jk} = 1$ and $0 \leq w_{jk} \leq 1$ can be solved with standard numerical optimization algorithms allowing for constraints, such as the L-BFGS-B method (Byrd et al., 1995). This algorithm is a limited memory modification of the quasi-Newton BFGS (Broyden-Fletcher-Goldfarb-Shanno) method, see e.g. Broyden (1970).

In practice we do not know the asymptotic variances and covariances $\mathsf{V}(\hat{\beta}_j(t_k))$ and $\mathsf{C}(\hat{\beta}_j(t_k), \hat{\beta}_j(t_l))$ in (4.10), they must be estimated from the data (Lemma 4.2). Replacing each $\mathsf{V}(\hat{\beta}_j(t_k))$ and $\mathsf{C}(\hat{\beta}_j(t_k), \hat{\beta}_j(t_l))$ by its estimator and solving the minimization problem leads to estimated optimal weights $\hat{w}_{j1}^{\text{opt}}, \dots, \hat{w}_{jK}^{\text{opt}}$ for each j. The resulting CLE with estimated weight matrices $\text{diag}\{\hat{\boldsymbol{w}}_1^{\text{opt}}\}, \dots, \text{diag}\{\hat{\boldsymbol{w}}_K^{\text{opt}}\}$ remains consistent and asymptotically normal (Theorem 4.6).

Combining estimators computed at failure times

In the previous part we proposed the combined logistic estimator for a known number K of fixed time points. To capture information from the proportional odds model we would like to perform the subanalyses as frequently as possible. Recall however that the likelihood function (4.4) of the proportional odds model can only change at times when one or more failures occur. Thus, the individual case-control analyses should be performed at all the observed failure times.

Combining the estimators computed at failure times has two major consequences. First, all results derived for fixed time points need to be adjusted for a random number of random time points and the number of analysis times K tends to infinity as the number of observations n increases. Nevertheless, the key results for random time points are all the same as for fixed time points (Theorem 4.12). The estimators remain consistent and asymptotically normal and we can apply the same principles for estimating the variances and covariances and computing optimal weights.

Second, with random times of the subanalyses we must assure that there will be enough cases for all subanalyses. Since we rely on asymptotic properties of the individual case-control estimators $\hat{\boldsymbol{\beta}}(t_k),\ k = 1, \dots, K$, assume there is a fixed starting time $\tau_0,\ \mathrm{P}(T < \tau_0) = c_0 > 0$ and include only subanalyses occurring after τ_0 in the linear combination. In practice we can eliminate the early estimators by choosing some threshold $k_0,\ 1 \leq k_0 < K$ and setting $w_{j1} = \cdots = w_{jk_0} = 0\ \forall j = 1, \dots, p$ (or shortly $\boldsymbol{W}_1 = \cdots = \boldsymbol{W}_{k_0} = \mathbf{0}_{(p \times p)}$). The optimization process is then only run for the remaining weight matrices $\boldsymbol{W}_{k_0+1}, \dots, \boldsymbol{W}_K$.

The practical choice of the threshold k_0 is a subjective decision. Smaller values of k_0 include more time information from the early estimators while larger values of k_0 discard more case-control analyses. Note that if we set $k_0 = K-1$, we obtain a simple case-control analysis equivalent to an analysis performed at the end of the study. We recommend to use a threshold $k_0 = \max\{30, k_\eta\}$, where $\lim_{K\to\infty} k_\eta/K = \eta > 0$. Other options for the threshold parameter are discussed later in Chapter 6.

Summary

We have proposed the combined logistic estimator for estimating regression parameters in the proportional odds model. With full data, the CLE is an alternative to the estimator proposed by Murphy et al. (1997). With case-cohort data, the CLE combines estimators from case-control logistic regression analyses. The combined logistic estimator thus does not use inverse sampling probabilities for weighting the individual contributions if applied to case-cohort data. The whole procedure works in the following five steps:

1. Choose the threshold index k_0.
2. Perform a case-control logistic regression analysis at all failure times and obtain the case-control estimates $\hat{\boldsymbol{\beta}}(t_k), k \geq k_0$.
3. Estimate the asymptotic covariance matrices for all $\hat{\boldsymbol{\beta}}(t_k)$ and the asymptotic covariances between the individual case-control estimators computed at different times, i.e. $\mathsf{C}(\hat{\boldsymbol{\beta}}(t_k), \hat{\boldsymbol{\beta}}(t_{k'})), k, k' = k_0, \ldots, K$.
4. Estimate the optimal weights and combine the case-control estimators into the combined logistic estimator $\tilde{\boldsymbol{\beta}}^{\text{opt}}$.
5. Estimate $\boldsymbol{\Sigma}_{\boldsymbol{W}^{\text{opt}}}$ – the asymptotic covariance matrix of $\tilde{\boldsymbol{\beta}}^{\text{opt}}$.

The resulting estimator $\tilde{\boldsymbol{\beta}}^{\text{opt}}$ is consistent, asymptotically normal and its components have the smallest variance among all convex linear combinations of logistic regression estimators. In particular, measured by the asymptotic variance of individual components of $\tilde{\boldsymbol{\beta}}^{\text{opt}}$, the combined logistic estimator is not worse than the last case-control logistic regression estimator $\hat{\boldsymbol{\beta}}(t_K)$.

4.3 Theoretical results

This section presents all theorems referenced in previous sections with proofs and technical details. We shall start with fixed time points and generalize the results step by step.

Notation

We assume that data are generated from the proportional odds model

$$\mathrm{logit}(\mathrm{P}(\delta(t) = 1|\boldsymbol{Z} = \boldsymbol{z})) = \alpha_0(t) + \boldsymbol{\beta}_0'\boldsymbol{z}, \tag{4.12}$$

where $\alpha_0(t)$ and $\boldsymbol{\beta}_0$ denote true parameter values. Parametrisation (4.12) is equivalent to the most common definition of the proportional odds model, which specifies $-\mathrm{logit}(S(t)|\boldsymbol{z}))$ on the left hand side (formula (2.6) in Chapter 2), because $S(t|\boldsymbol{z}) = 1 - \mathrm{P}(\delta(t) = 1|\boldsymbol{z})$ and the minus sign reverses the logit. It follows that

$$\mathrm{P}(\delta(t) = 1|\boldsymbol{Z} = \boldsymbol{z}) = \frac{\exp\{\alpha_0(t) + \boldsymbol{\beta}_0'\boldsymbol{z}\}}{1 + \exp\{\alpha_0(t) + \boldsymbol{\beta}_0'\boldsymbol{z}\}}. \tag{4.13}$$

Denote the true conditional probability of an event in (4.13) by $\pi_0(t, \boldsymbol{z})$ and define $\pi(\cdot)$ as a function of time, parameters and covariates

$$\pi(t) = \pi(t, \alpha, \boldsymbol{\beta}, \boldsymbol{z}) = \frac{\exp\{\alpha(t) + \boldsymbol{\beta}'\boldsymbol{z}\}}{1 + \exp\{\alpha(t) + \boldsymbol{\beta}'\boldsymbol{z}\}}. \tag{4.14}$$

Note that (4.14) no longer has the interpretation of event probability unless evaluated at the true parameters $\alpha_0(t)$ and $\boldsymbol{\beta}_0$, then we write $\pi_0(t)$. In case we need to specify the dependence on a particular set of covariates $\boldsymbol{z}_i$ for individual i, we shall use $\pi_i(t)$ or $\pi_{i0}(t)$.

For $t > \tau_0$ denote by $\hat{\boldsymbol{\theta}}(t) = (\hat{\alpha}(t), \hat{\boldsymbol{\beta}}(t))'$ the maximum likelihood estimators obtained from a logistic regression model fitted to data at time t. Such analysis is retrospective, based on case-control data, and according to Prentice & Pyke (1979), $\hat{\alpha}(t)$ and $\hat{\boldsymbol{\beta}}(t)$ estimate parameters $\alpha_0^\star(t)$ and $\boldsymbol{\beta}_0$. With retrospective data, $\alpha_0^\star(t)$ is generally different from the true value $\alpha_0(t)$. As summarised in Section 3.5, for each t the estimator $\hat{\boldsymbol{\beta}}(t)$ remains consistent, asymptotically normal and its variance has the same form as in a prospective study. Results concerning the intercept are affected by the retrospective case-control design, so we need to treat the parameters separately and distinguish between the intercept term $\alpha(t)$ and the $p \times 1$ vector $\boldsymbol{\beta}$. Similarly as in (4.14) we denote the case-control event probability by $\pi_0^\star(t) = \pi(t, \alpha_0^\star(t), \boldsymbol{\beta}_0, \boldsymbol{z})$, its estimator $\hat{\pi}(t) = \pi(t, \hat{\alpha}(t), \hat{\boldsymbol{\beta}}(t), \boldsymbol{z})$ and use i in the subscript for expressing the dependence on $\boldsymbol{z}_i$.

Assumptions

Throughout the whole section on theoretical results we assume that

i Censoring can only occur at the end of the study

ii All covariates are bounded.

We will discuss relaxing or weakening of the conditions later in Chapter 6.

CLE with fixed times of individual analyses

Assume for now that there are K fixed time points $t_1 < t_2 < \cdots < t_K$ at which the study is stopped and analysed by a logistic regression model.

Lemma 4.1: *For each $s \in \{t_1, \ldots, t_K\}$ the normalized logistic regression score function evaluated at $(\alpha_0^\star(s), \boldsymbol{\beta}_0)^T$ is asymptotically normally distributed*

$$\frac{1}{\sqrt{n}} \begin{pmatrix} U_\alpha(s, \alpha_0^\star(s)) \\ \boldsymbol{U}_{\boldsymbol{\beta}}(s, \boldsymbol{\beta}_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{U}}(s)),$$

where the elements of the asymptotic variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{U}}(s)$ equal

$$\begin{aligned} \mathsf{C}_{\alpha,\alpha}(s) &= \mathsf{E}_{\boldsymbol{Z}}\{\pi_0^\star(s)[1-\pi_0^\star(s)]\} \\ \mathsf{C}_{\alpha,\beta_j}(s) &= \mathsf{E}_{\boldsymbol{Z}}\{Z_j \pi_0^\star(s)[1-\pi_0^\star(s)]\}, \ j = 1, \ldots, p \\ \mathsf{C}_{\beta_j,\beta_{j'}}(s) &= \mathsf{E}_{\boldsymbol{Z}}\{Z_j Z_{j'} \pi_0^\star(s)[1-\pi_0^\star(s)]\}, \ j, j' = 1, \ldots, p. \end{aligned} \tag{4.15}$$

The asymptotic variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{U}}(s)$ can be consistently estimated by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(s)$ with elements

$$\begin{aligned} \hat{\mathsf{C}}_{\alpha,\alpha}(s) &= \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}_i(s)[1-\hat{\pi}_i(s)] \\ \hat{\mathsf{C}}_{\alpha,\beta_j}(s) &= \frac{1}{n}\sum_{i=1}^{n} z_{i,j}\hat{\pi}_i(s)[1-\hat{\pi}_i(s)], \ j = 1, \ldots, p \\ \hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(s) &= \frac{1}{n}\sum_{i=1}^{n} z_{i,j} z_{i,j'}\hat{\pi}_i(s)[1-\hat{\pi}_i(s)], \ j, j' = 1, \ldots, p. \end{aligned}$$

Proof: For full prospective data, Lemma 4.1 with $\pi_0^\star(\cdot)$ replaced by $\pi_0(\cdot)$ is a standard result from theory of maximum likelihood estimation. Regularity conditions are satisfied for logistic regression scores and the asymptotic normality follows from the Central limit theorem for independent identically distributed random variables. The form of the asymptotic variance matrix and its estimator can be found in standard textbooks. The generalisation for case-control studies follows from Prentice & Pyke (1979). ■

Lemma 4.2: *The vector of normalized logistic regression score functions calculated at times $t_1, \ldots, t_K$ is jointly asymptotically normally distributed*

$$n^{-1/2} \begin{pmatrix} U_\alpha(t_1, \alpha_0^\star(t_1)) \\ \boldsymbol{U}_{\boldsymbol{\beta}}(t_1, \boldsymbol{\beta}_0) \\ \vdots \\ U_\alpha(t_K, \alpha_0^\star(t_K)) \\ \boldsymbol{U}_{\boldsymbol{\beta}}(t_K, \boldsymbol{\beta}_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1, \ldots, t_K)),$$

where the asymptotic variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1,\ldots,t_K)$ *is a* $K(p+1)\times K(p+1)$ *matrix composed of* K^2 *blocks* $\boldsymbol{\Sigma}_{\boldsymbol{U}}(s,t)$ *with* $s=t_k,\ t=t_{k'},\ k,k'=1,\ldots,K,\ s\le t$. *The elements of each* $\boldsymbol{\Sigma}_{\boldsymbol{U}}(s,t)$ *are given by*

$$\begin{aligned}
\mathsf{C}_{\alpha,\alpha}(s,t) &= \mathsf{E}_{\boldsymbol{Z}}\{\pi_0^\star(s)[1-\pi_0^\star(t)]\}\\
\mathsf{C}_{\alpha,\beta_j}(s,t) &= \mathsf{E}_{\boldsymbol{Z}}\{Z_j\pi_0^\star(s)[1-\pi_0^\star(t)]\},\ j=1,\ldots,p\\
\mathsf{C}_{\beta_j,\beta_{j'}}(s,t) &= \mathsf{E}_{\boldsymbol{Z}}\{Z_jZ_{j'}\pi_0^\star(s)[1-\pi_0^\star(t)]\},\ j,j'=1,\ldots,p.
\end{aligned}$$

Note that the off-diagonal blocks of $\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1,\ldots,t_K)$ contain asymptotic covariances of normalised scores computed at different time points, while its diagonal blocks $\boldsymbol{\Sigma}_{\boldsymbol{U}}(s,s)=\boldsymbol{\Sigma}_{\boldsymbol{U}}(s)$ are the asymptotic variance matrices of the individual normalised logistic regression scores from Lemma 4.1.

The asymptotic marginal distribution of scores for $\boldsymbol{\beta}$ alone (without α scores) is also multivariate normal with zero mean. Its asymptotic variance matrix is composed of appropriate rows and columns from $\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1,\ldots,t_K)$.

Proof: The asymptotic normality in Lemma 4.2 follows from the multivariate Central limit theorem for independent identically distributed random variables since we can write

$$\begin{pmatrix} U_\alpha(t_1,\alpha_0^\star(t_1))\\ \boldsymbol{U}_{\boldsymbol{\beta}}(t_1,\boldsymbol{\beta}_0)\\ \vdots\\ U_\alpha(t_K,\alpha_0^\star(t_K))\\ \boldsymbol{U}_{\boldsymbol{\beta}}(t_K,\boldsymbol{\beta}_0)\end{pmatrix}=\sum_{i=1}^n\begin{pmatrix} U_{i,\alpha}(t_1,\alpha_0^\star(t_1))\\ \boldsymbol{U}_{i,\boldsymbol{\beta}}(t_1,\boldsymbol{\beta}_0)\\ \vdots\\ U_{i,\alpha}(t_K,\alpha_0^\star(t_K))\\ \boldsymbol{U}_{i,\boldsymbol{\beta}}(t_K,\boldsymbol{\beta}_0)\end{pmatrix}=\sum_{i=1}^n\begin{pmatrix}\delta_i(t_1)-\pi_{0,i}^\star(t_1)\\ Z_{i,1}[\delta_i(t_1)-\pi_{0,i}^\star(t_1)]\\ \vdots\\ Z_{i,p}[\delta_i(t_1)-\pi_{0,i}^\star(t_1)]\\ \vdots\\ \delta_i(t_K)-\pi_{0,i}^\star(t_K)\\ Z_{i,1}[\delta_i(t_K)-\pi_{0,i}^\star(t_K)]\\ \vdots\\ Z_{i,p}[\delta_i(t_K)-\pi_{0,i}^\star(t_K)]\end{pmatrix}.$$

It remains to derive the blocks of the asymptotic variance matrix. Denote by $\boldsymbol{U}_i$ the score contribution from the i^{th} individual and recall that with full (prospective) data, $\delta_i(t)$ is a binary random variable with conditional expectation $\mathsf{E}[\delta_i(t)|\boldsymbol{z}]=\pi_0(t)$ and variance $\mathsf{Var}[\delta_i(t)|\boldsymbol{z}]=\pi_0(t)[1-\pi_0(t)]$ for each $i=1,\ldots,n$. In a case-control study we have $\mathsf{E}[\delta_i(t)|\boldsymbol{z}]=\pi_0^\star(t)$, $\mathsf{Var}[\delta_i(t)|\boldsymbol{z}]=\pi_0^\star(t)[1-\pi_0^\star(t)]$ and we can write

$$\begin{aligned}
\mathsf{Cov}[U_{i,\alpha}(s),U_{i,\alpha}(t)|\boldsymbol{z}] =&\ \mathsf{Cov}[U_{i,\alpha}(s),U_{i,\alpha}(s)|\boldsymbol{z}]\\
&+\mathsf{Cov}[U_{i,\alpha}(s),(U_{i,\alpha}(t)-U_{i,\alpha}(s))|\boldsymbol{z}]\\
=&\ \mathsf{Var}[U_{i,\alpha}(s)|\boldsymbol{z}]+\mathsf{E}\{U_{i,\alpha}(s)[U_{i,\alpha}(t)-U_{i,\alpha}(s)]|\boldsymbol{z}\},
\end{aligned}$$

since $\mathsf{E}[U_{i,\alpha}(s)|\boldsymbol{z}]=0$. Therefore

$$\begin{aligned}\mathsf{Cov}[U_{i,\alpha}(s), U_{i,\alpha}(t)|\boldsymbol{z}] =& \ \mathsf{V}\mathrm{ar}[\delta_i(s) - \pi_0^\star(s)|\boldsymbol{z}]\\ &+ \mathsf{E}\{[\delta_i(s) - \pi_0^\star(s)][\delta_i(t) - \pi_0^\star(t) - \delta_i(s) + \pi_0^\star(s)]|\boldsymbol{z}\}\\ =& \ \pi_0^\star(s)[1 - \pi_0^\star(s)]\\ &+ \mathsf{E}\{\delta_i(s)[\delta_i(t) - \delta_i(s) - (\pi_0^\star(t) - \pi_0^\star(s))]|\boldsymbol{z}\}\\ &- \mathsf{E}\{\pi_0^\star(s)[\delta_i(t) - \delta_i(s) - (\pi_0^\star(t) - \pi_0^\star(s))]|\boldsymbol{z}\}.\end{aligned}$$

Now, the last term above equals 0 while from the previous one we obtain

$$\begin{aligned}\mathsf{E}\{\delta_i(s)[\pi_0^\star(t) - \pi_0^\star(s)]|\boldsymbol{z}\} &= \pi_0^\star(s)(\pi_0^\star(t) - \pi_0^\star(s))\\ \text{and } \ \mathsf{E}\{\delta_i(s)[\delta_i(t) - \delta_i(s)]|\boldsymbol{z}\} &= 0,\end{aligned}$$

because the product $\delta_i(s)[\delta_i(t) - \delta_i(s)]$ is constantly 0. Putting all together yields

$$\begin{aligned}\mathsf{Cov}[U_{i,\alpha}(s), U_{i,\alpha}(t)|\boldsymbol{z}] &= \pi_0^\star(s)[1 - \pi_0^\star(s)] - \pi_0^\star(s)[\pi_0^\star(t) - \pi_0^\star(s)]\\ &= \pi_0^\star(s)[1 - \pi_0^\star(t)]\end{aligned}$$

implying the unconditional covariance

$$\begin{aligned}\mathsf{Cov}[U_{i,\alpha}(s), U_{i,\alpha}(t)] =& \ \mathsf{E}_{\boldsymbol{Z}}\{\mathsf{Cov}[U_{i,\alpha}(s), U_{i,\alpha}(t)|\boldsymbol{z}]\}\\ =& \ \mathsf{E}_{\boldsymbol{Z}}\, \pi_0^\star(s)[1 - \pi_0^\star(t)].\end{aligned}$$

The remaining covariances involving scores for components of $\boldsymbol{\beta}$ can be calculated similarly. ■

Lemma 4.3: *The asymptotic variance matrix of scores $\boldsymbol{\Sigma_U}(t_1, \ldots, t_K)$ introduced in Lemma 4.2 can be consistently estimated by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_1, \ldots, t_K)$, a matrix composed of K^2 blocks $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(s,t)$, $s = t_k$, $t = t_{k'}$, $k, k' = 1, \ldots, K$ with elements*

$$\begin{aligned}\hat{\mathsf{C}}_{\alpha,\alpha}(s,t) &= \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}_i(s)[1 - \hat{\pi}_i(t)]\\ \hat{\mathsf{C}}_{\alpha,\beta_j}(s,t) &= \frac{1}{n}\sum_{i=1}^{n} z_{i,j}\hat{\pi}_i(s)[1 - \hat{\pi}_i(t)], \ j = 1, \ldots, p\\ \hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(s,t) &= \frac{1}{n}\sum_{i=1}^{n} z_{i,j}z_{i,j'}\hat{\pi}_i(s)[1 - \hat{\pi}_i(t)], \ j, j' = 1, \ldots, p.\end{aligned}$$

Proof: We shall only present the proof for $\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(s,t)$ in detail, the proof for other elements of $\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1,\dots,t_K)$ is analogous. The estimated parameters $\hat{\alpha}(s)$ and $\hat{\boldsymbol{\beta}}(s)$ are contained in $\hat{\pi}(s)$, which makes the summands dependent. Standard solution is to rewrite each term $\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(s,t)$, which should be in fact denoted by $\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\hat{\alpha}(s),\hat{\boldsymbol{\beta}}(s),\hat{\alpha}(t),\hat{\boldsymbol{\beta}}(t))$ to reflect its dependence on parameters, as $\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\alpha_0^\star(s),\boldsymbol{\beta}_0,\alpha_0^\star(t),\boldsymbol{\beta}_0)$ plus an asymptotically negligible remainder term. First order Taylor expansion of $\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\hat{\alpha}(s),\hat{\boldsymbol{\beta}}(t),\hat{\alpha}(t),\hat{\boldsymbol{\beta}}(t))$ around $(\alpha_0^\star(s),\boldsymbol{\beta}_0,\alpha_0^\star(t),\boldsymbol{\beta}_0)$ gives

$$\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\hat{\alpha}(s),\hat{\boldsymbol{\beta}}(s),\hat{\alpha}(t),\hat{\boldsymbol{\beta}}(t)) \doteq \hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\alpha_0^\star(s),\boldsymbol{\beta}_0,\alpha_0^\star(t),\boldsymbol{\beta}_0) + \begin{pmatrix} \hat{\alpha}(s)-\alpha_0^\star(s) \\ \hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0 \\ \hat{\alpha}(t)-\alpha_0^\star(t) \\ \hat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0 \end{pmatrix}^T \begin{pmatrix} \frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\alpha}(s)}\Big|_{\alpha_0^\star(s),\boldsymbol{\beta}_0} \\ \frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\boldsymbol{\beta}}(s)}\Big|_{\alpha_0^\star,\boldsymbol{\beta}_0} \\ \frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\alpha}(t)}\Big|_{\alpha_0^\star(t),\boldsymbol{\beta}_0} \\ \frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\boldsymbol{\beta}}(t)}\Big|_{\alpha_0^\star,\boldsymbol{\beta}_0} \end{pmatrix}$$

where

$$\begin{aligned}
\frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\alpha}(s)}\Big|_{\alpha_0^\star(s),\boldsymbol{\beta}_0} &= \frac{1}{n}\sum_{i=1}^n \frac{\partial z_{i,j}z_{i,j'}\pi_i^\star(s)(1-\pi_i^\star(t))}{\partial\hat{\alpha}(s)} \\
&= \frac{1}{n}\sum_{i=1}^n \frac{z_{i,j}z_{i,j'}}{1+\exp(\alpha_0^\star(s)+\boldsymbol{\beta}_0'\boldsymbol{z}_i)}\pi_i^\star(s)(1-\pi_i^\star(t)), \\
\frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\alpha}(t)}\Big|_{\alpha_0^\star(t),\boldsymbol{\beta}_0} &= -\frac{1}{n}\sum_{i=1}^n \frac{z_{i,j}z_{i,j'}}{1+\exp(\alpha_0^\star(t)+\boldsymbol{\beta}_0'\boldsymbol{z}_i)}\pi_i^\star(s)(1-\pi_i^\star(t)), \qquad (4.16) \\
\frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\beta}_l(s)}\Big|_{\alpha_0^\star(s),\boldsymbol{\beta}_0} &= \frac{1}{n}\sum_{i=1}^n \frac{z_{ij}z_{i,j'}z_{i,l}}{1+\exp(\alpha_0^\star(s)+\boldsymbol{\beta}_0'\boldsymbol{z}_i)}\pi_i^\star(s)(1-\pi_i^\star(t)), \\
\frac{\partial\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}}{\partial\hat{\beta}_l(t)}\Big|_{\alpha_0^\star(t),\boldsymbol{\beta}_0} &= -\frac{1}{n}\sum_{i=1}^n \frac{z_{ij}z_{i,j'}z_{i,l}}{1+\exp(\alpha_0^\star(t)+\boldsymbol{\beta}_0'\boldsymbol{z}_i)}\pi_i^\star(s)(1-\pi_i^\star(t))
\end{aligned}$$

for $l=1,\dots,p$. Now, $\{(\hat{\alpha}(s),\hat{\boldsymbol{\beta}}(s),\hat{\alpha}(t),\hat{\boldsymbol{\beta}}(t))-(\alpha_0^\star(s),\boldsymbol{\beta}_0,\alpha_0^\star(t),\boldsymbol{\beta}_0)\}$ tends to $\mathbf{0}$ in probability (consistency of the estimators) and the derivatives are bounded in probability due to assumptions imposed on the covariates. Thus

$$\hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\hat{\alpha}(s),\hat{\boldsymbol{\beta}}(s),\hat{\alpha}(t),\hat{\boldsymbol{\beta}}(t)) = \hat{\mathsf{C}}_{\beta_j,\beta_{j'}}(\alpha_0^\star(s),\boldsymbol{\beta}_0,\alpha_0^\star(t),\boldsymbol{\beta}_0)+o_P(1),$$

terms on the right hand side are independent identically distributed and the law of large numbers applies to complete the proof. ■

Theorem 4.4: *The vector of normalized logistic regression estimators computed at times $t_1, \ldots, t_K$ is jointly asymptotically normally distributed*

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}(t_1) - \alpha_0^\star(t_1) \\ \hat{\boldsymbol{\beta}}(t_1) - \boldsymbol{\beta}_0 \\ \vdots \\ \hat{\alpha}(t_K) - \alpha_0^\star(t_K) \\ \hat{\boldsymbol{\beta}}(t_K) - \boldsymbol{\beta}_0 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(t_1, \ldots, t_K)),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(t_1, \ldots, t_K) = \boldsymbol{J}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1, \ldots, t_K)\boldsymbol{J}^{-1}$ is a $K(p+1) \times K(p+1)$ matrix, $\boldsymbol{J}$ is a block-diagonal matrix $\mathrm{diag}\{\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1), \ldots, \boldsymbol{\Sigma}_{\boldsymbol{U}}(t_K)\}$, $\boldsymbol{\Sigma}_{\boldsymbol{U}}(s)$ is the asymptotic variance matrix of logistic score at time s (Lemma 4.2) and $\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1, \ldots, t_K)$ is the asymptotic variance matrix of logistic scores computed at times $t_1, \ldots, t_K$ (Lemma 4.3).

Note in particular, that Theorem 4.4 implies joint asymptotic normality

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}(t_1) - \boldsymbol{\beta}_0 \\ \vdots \\ \hat{\boldsymbol{\beta}}(t_K) - \boldsymbol{\beta}_0 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1, \ldots, t_K)),$$

where the asymptotic $Kp \times Kp$ variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1, \ldots, t_K)$ consists of appropriate rows and columns of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(t_1, \ldots, t_K)$.

Since $\boldsymbol{J}$ is a block-diagonal matrix, its inverse $\boldsymbol{J}^{-1}$ is also a block-diagonal matrix, $\boldsymbol{J}^{-1} = \mathrm{diag}\{\{\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1)\}^{-1}, \ldots, \{\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_K)\}^{-1}\}$. Splitting the expression $\boldsymbol{J}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1, \ldots, t_K)\boldsymbol{J}^{-1}$ according to diagonal blocks of $\boldsymbol{J}^{-1}$ provides for each t_k the usual result for logistic regression estimators.

Proof: It is sufficient to provide a proof for $K = 2$ and $p = 1$, that is two logistic regression estimators and a single covariate. The extension to any finite number K and more covariates is straightforward. From Lemma 4.2 we can see that

$$n^{-1/2}\begin{pmatrix} U_\alpha(t_1, \alpha_0^\star(t_1)) \\ U_\beta(t_1, \beta_0) \\ U_\alpha(t_2, \alpha_0^\star(t_2)) \\ U_\beta(t_2, \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1, t_2)).$$

Using first order Taylor expansion of $(\boldsymbol{U}(t_1, \hat{\alpha}(t_1), \hat{\beta}(t_1)), \boldsymbol{U}(t_2, \hat{\alpha}(t_2), \hat{\beta}(t_2)))^T$ around $(\alpha_0^\star(t_1), \beta_0, \alpha_0^\star(t_2), \beta_0)^T$ together with Lemma 4.3 gives

$$\frac{1}{\sqrt{n}}\begin{pmatrix} U_\alpha(t_1,\hat{\alpha}(t_1),\hat{\beta}(t_1)) \\ U_\beta(t_1,\hat{\alpha}(t_1),\hat{\beta}(t_1)) \\ U_\alpha(t_2,\hat{\alpha}(t_2),\hat{\beta}(t_2)) \\ U_\beta(t_2,\hat{\alpha}(t_2),\hat{\beta}(t_2)) \end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix} U_\alpha(t_1,\alpha_0^\star(t_1),\boldsymbol{\beta}_0) \\ U_\beta(t_1,\alpha_0^\star(t_1),\boldsymbol{\beta}_0) \\ U_\alpha(t_2,\alpha_0^\star(t_2),\boldsymbol{\beta}_0) \\ U_\beta(t_2,\alpha_0^\star(t_2),\boldsymbol{\beta}_0) \end{pmatrix}$$
$$- \boldsymbol{J}\,\sqrt{n}\begin{pmatrix} \hat{\alpha}(t_1)-\alpha_0^\star(t_1) \\ \hat{\beta}(t_1)-\beta_0 \\ \hat{\alpha}(t_2)-\alpha_0^\star(t_2) \\ \hat{\beta}(t_2)-\beta_0 \end{pmatrix} + o_P(1) = \mathbf{0},$$

leading directly to

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}(t_1)-\alpha_0^\star(t_1) \\ \hat{\beta}(t_1)-\boldsymbol{\beta}_0 \\ \hat{\alpha}(t_2)-\alpha_0^\star(t_2) \\ \hat{\beta}(t_2)-\boldsymbol{\beta}_0 \end{pmatrix} = \frac{1}{\sqrt{n}}\boldsymbol{J}^{-1}\begin{pmatrix} U_\alpha(t_1,\alpha_0^\star(t_1),\beta_0) \\ U_\beta(t_1,\alpha_0^\star(t_1),\beta_0) \\ U_\alpha(t_2,\alpha_0^\star(t_2),\beta_0) \\ U_\beta(t_2,\alpha_0^\star(t_2),\beta_0) \end{pmatrix} + o_P(1). \quad (4.17)$$

The scores are asymptotically jointly normal, thus the estimators are also asymptotically jointly normal and the proof is complete. ■

The joint multivariate normality implies that also all subvectors and their linear combinations are asymptotically normal. Theorem 4.4 therefore serves as a basis for establishing asymptotic normality and the asymptotic variance matrix for the class of combined logistic estimators $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}$.

Corollary 4.5: *Any estimator $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}$ belonging to the class of combined logistic estimators defined in* (4.7)*, and in particular the optimal combined logistic estimator $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}^{\text{opt}}}$ defined on page 44, is asymptotically normal*

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{W}(t_1,\ldots,t_K)}),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{W}(t_1,\ldots,t_K)} = \sum_{k=1}^{K}\sum_{k'=1}^{K} \boldsymbol{W}_k\, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_k,t_{k'})\boldsymbol{W}_{k'}{}^T. \quad (4.18)$$

The asymptotic covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_k,t_{k'})$ are $p\times p$ blocks taken from the asymptotic variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1,\ldots,t_K)$. ■

Theorem 4.6: *The asymptotic variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(t_1,\ldots,t_K)$ from Theorem 4.4 can be consistently estimated by*

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}(t_1,\ldots,t_K) = \hat{\boldsymbol{J}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_1,\ldots,t_K)\hat{\boldsymbol{J}}^{-1}, \quad (4.19)$$

where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_1, \ldots, t_K)$ *is the estimated asymptotic variance matrix of normalised scores computed at* $t_1, \ldots, t_K$*, see Lemma 4.3, and* $\hat{\boldsymbol{J}}$ *is the estimated block-diagonal matrix* $\hat{\boldsymbol{J}} = \mathrm{diag}\{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_1), \ldots, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_K)\}$.

Proof: A similar result was shown for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_1, \ldots, t_K)$ in Lemma 4.3, the proof for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}(t_1, \ldots, t_K)$ proceeds in the same way. ■

Theorem 4.6 shows that if we replace the unknown values of $\alpha_0^\star(s)$ and $\boldsymbol{\beta}_0$ by their estimated versions $\hat{\alpha}(s)$ and $\hat{\boldsymbol{\beta}}(s)$ for $s \in \{t_1, \ldots, t_K\}$, we obtain a consistent estimator of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}(t_1, \ldots, t_K)$ and consequently also for its submatrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_1, \ldots, t_K)$. Similarly, if the unknown true covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_k, t_{k'})$ are replaced by their consistent estimators $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_k, t_{k'})$ in (4.18), we obtain $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}, \boldsymbol{W}(t_1, \ldots, t_K)}$, a consistent estimator of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}, \boldsymbol{W}(t_1, \ldots, t_K)}$ for any fixed weight matrices $\boldsymbol{W}(t_1, \ldots, t_K)$.

However, when computing the optimal combined estimator, we are looking for weights that minimize the variance of individual components of $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}$. Therefore the optimal weights are some function of the asymptotic variances and covariances, they are random and need to be estimated. The following theorem assures that we can use the estimated weights.

Theorem 4.7: *Let the optimal weights be defined as* $\boldsymbol{W}^{\mathrm{opt}}(t_1, \ldots, t_K) = \boldsymbol{\phi}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1, \ldots, t_K))$ *for some continuous function* $\boldsymbol{\phi}$*. Define the estimated weights by plugging in the estimated variance matrix,* $\hat{\boldsymbol{W}}^{\mathrm{opt}}(t_1, \ldots, t_K) = \boldsymbol{\phi}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_1, \ldots, t_K))$*. Then*

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\hat{\boldsymbol{W}}^{\mathrm{opt}}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}, \boldsymbol{W}^{\mathrm{opt}}(t_1, \ldots, t_K)}),$$

where $\tilde{\boldsymbol{\beta}}_{\hat{\boldsymbol{W}}^{\mathrm{opt}}} = \sum_{k=1}^{K} \hat{\boldsymbol{W}}_k^{\mathrm{opt}} \hat{\boldsymbol{\beta}}(t_k)$ *is the optimal combined logistic estimator computed using the estimated weights.*

Proof: Both the true as well as estimated weights must add up to the identity matrix: $\sum_k \hat{\boldsymbol{W}}_k = \sum_k \boldsymbol{W}_k = \boldsymbol{I}_{p \times p}$. Therefore

$$\begin{aligned}
\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\hat{\boldsymbol{W}}^{\mathrm{opt}}} - \boldsymbol{\beta}_0) &= \sqrt{n}\left(\sum_{k=1}^{K} \hat{\boldsymbol{W}}_k^{\mathrm{opt}} \hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0\right) \\
&= \sum_{k=1}^{K} \hat{\boldsymbol{W}}_k^{\mathrm{opt}} \sqrt{n}\left(\hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0\right) \\
&= \sum_{k=1}^{K} \boldsymbol{W}_k^{\mathrm{opt}} \sqrt{n}\left(\hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0\right) \\
&\quad + \sum_{k=1}^{K} \left(\hat{\boldsymbol{W}}_k^{\mathrm{opt}} - \boldsymbol{W}_k^{\mathrm{opt}}\right) \sqrt{n}\left(\hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{n}\left(\sum_{k=1}^{K} \boldsymbol{W}_k^{\mathrm{opt}} \hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0\right) \\
&\quad + \sum_{k=1}^{K} \Big(\boldsymbol{\phi}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_1,\dots,t_K)) - \boldsymbol{\phi}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1,\dots,t_K))\Big) \\
&\quad \times \sqrt{n}\left(\hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0\right) \\
&= \sqrt{n}\left(\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}^{\mathrm{opt}}} - \boldsymbol{\beta}_0\right) + o_P(1),
\end{aligned}
$$

since $\boldsymbol{\phi}$ is continuous, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_1,\dots,t_K)$ and $\hat{\boldsymbol{\beta}}(t_k)$ are consistent estimators of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1,\dots,t_K)$ and $\boldsymbol{\beta}_0$, respectively, and $\sqrt{n}(\hat{\boldsymbol{\beta}}(t_k) - \boldsymbol{\beta}_0)$ is asymptotically normal for all $k = 1,\dots,K$. The rest follows from Corollary 4.5. ■

Finally, if we use the estimated weights and estimated covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_k, t_{k'})$, we obtain a consistent estimator of the variance-covariance matrix of the optimal combined logistic estimator. The elements of $\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{W}^{\mathrm{opt}}(t_1,\dots,t_K)}$ for fixed weights are given by formulas (4.10), the estimator $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},\hat{\boldsymbol{W}}^{\mathrm{opt}}(t_1,\dots,t_K)}$ consists of

$$
\begin{aligned}
\hat{\mathsf{V}}(\tilde{\beta}_j) &= \sum_{k=1}^{K} \hat{w}_{jk}^2 \hat{\mathsf{V}}(\hat{\beta}_{jk}) + 2\sum_{k=1}^{K}\sum_{l=k+1}^{K} \hat{w}_{jk}\hat{w}_{jl}\hat{\mathsf{C}}(\hat{\beta}_{jk}, \hat{\beta}_{jl}), \\
\hat{\mathsf{C}}(\tilde{\beta}_j, \tilde{\beta}_{j'}) &= \sum_{k=1}^{K} \hat{w}_{jk}\hat{w}_{j'k}\hat{\mathsf{C}}(\hat{\beta}_{jk}, \hat{\beta}_{j'k}) + \sum_{k=1}^{K}\sum_{l=1,\, l\neq k}^{K} \hat{w}_{jk}\hat{w}_{j'l}\hat{\mathsf{C}}(\hat{\beta}_{jk}, \hat{\beta}_{j'l}).
\end{aligned}
\tag{4.20}
$$

Consistency of the estimators defined in (4.20) is obvious.

Theorems 4.4 – 4.7 justify the intuitive development of the combined logistic estimator computed from a finite number of fixed time points. Now we need to generalise the results for combined logistic estimators assembled from individual analyses performed at random times, particularly at the failure times.

CLE with individual analyses performed at failure times

Consider now $\boldsymbol{U}(t)$, $\hat{\alpha}(t)$ and $\hat{\boldsymbol{\beta}}(t)$ as random processes that generalize the traditional logistic regression score function and parameter estimators. For convenience denote the whole parameter process by $\boldsymbol{\theta}(t) = (\alpha(t), \boldsymbol{\beta}(t))'$, by $\hat{\boldsymbol{\theta}}(t) = (\hat{\alpha}(t), \hat{\boldsymbol{\beta}}(t))'$ the estimator and by $\boldsymbol{\theta}_0(t) = (\alpha_0^\star(t), \boldsymbol{\beta}_0)'$ true parameter values. Let us now restate an important result of Kulich & Lin (2004), which will be used as a basis for deriving weak convergence of $\boldsymbol{U}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ to Gaussian processes.

Lemma 4.8: *Let $B_i(t), i = 1, \ldots, n$ be independent and identically distributed real-valued random processes on $[0, \tau]$ with $\mathsf{E}\, B_i(t) = \mu_B(t)$ and finite variances $\mathsf{Var}\, B_i(0) < \infty$ and $\mathsf{Var}\, B_i(\tau) < \infty$. Suppose that almost all paths of $B_i(t)$ have finite variation. Then $n^{-1/2} \sum_i [B_i(t) - \mu_B(t)]$ converges weakly in $\ell^\infty[0, \tau]$ to a zero-mean Gaussian process, and $n^{-1} \sum_i B_i(t)$ converges in probability to $\mu_B(t)$ uniformly in t.*

Proof: The proof is given in the Appendix of Kulich & Lin (2004), Proposition A1. ■

Theorem 4.9: *Let the baseline log odds function $\alpha_0^\star(t)$ have finite variation. The estimator $\hat{\boldsymbol{\theta}}(t)$ is consistent uniformly in t, i.e.*

$$\sup_{\tau_0 \le t \le \tau} \left|\left| \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t) \right|\right| \xrightarrow{\mathrm{P}} 0. \tag{4.21}$$

Proof: Using first order Taylor expansion of $\frac{1}{n}\boldsymbol{U}(t, \hat{\boldsymbol{\theta}}(t))$ around $\boldsymbol{\theta}_0(t)$ gives

$$\mathbf{0} = \frac{1}{n}\boldsymbol{U}(t, \hat{\boldsymbol{\theta}}(t)) = \frac{1}{n}\boldsymbol{U}(t, \boldsymbol{\theta}_0(t)) + \frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{\theta}(t)}\right|_{\boldsymbol{\theta}(t)=\boldsymbol{\theta}^\bullet(t)} \Big(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t)\Big)$$

for some $\boldsymbol{\theta}^\bullet(t)$ between $\hat{\boldsymbol{\theta}}(t)$ and $\boldsymbol{\theta}_0(t)$. Therefore

$$\begin{aligned}
&\sup_{\tau_0 \le t \le \tau} \left|\left| \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t) \right|\right| \\
&\quad = \sup_{\tau_0 \le t \le \tau} \left|\left| \left[\frac{1}{n}\sum_{i=1}^{n} - \left.\frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{\theta}(t)}\right|_{\boldsymbol{\theta}(t)=\boldsymbol{\theta}^\bullet(t)} \right]^{-1} \frac{1}{n}\boldsymbol{U}(t, \boldsymbol{\theta}_0(t)) \right|\right| \\
&\quad \le \sup_{\tau_0 \le t \le \tau} \left|\left| \boldsymbol{J}^{-1}(t) \frac{1}{n}\boldsymbol{U}(t, \boldsymbol{\theta}_0(t)) \right|\right| \\
&\qquad + \sup_{\tau_0 \le t \le \tau} \left|\left| \left[\left[\frac{1}{n}\sum_{i=1}^{n} - \left.\frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{\theta}(t)}\right|_{\boldsymbol{\theta}(t)=\boldsymbol{\theta}^\bullet(t)} \right]^{-1} - \boldsymbol{J}^{-1}(t) \right] \frac{1}{n}\boldsymbol{U}(t, \boldsymbol{\theta}_0(t)) \right|\right|,
\end{aligned}$$

where $\boldsymbol{J}(t)$ is the expected information matrix for $\boldsymbol{\theta}$ at time t. Now we need to verify the assumptions of Lemma 4.8 for the score process and its derivative to show that the right hand side tends to $\mathbf{0}$.

The expectation of $\boldsymbol{U}_i(t)$ equals $\mathbf{0}$ for all t, since $\boldsymbol{U}_i(t)$ is a piecewise constant process which can only jump at times of individual analyses t_k and at each t_k, $\boldsymbol{U}_i(t_k)$ is a contribution to the case-control logistic regression score with mean zero. For the variance we have that

$$\begin{aligned}\mathsf{Var}\, U_{i,\alpha}(t) =&\ \mathsf{E}_{\boldsymbol{Z}}\, \pi_0^\star(t)(1-\pi_0^\star(t))\\ \mathsf{Var}\, U_{i,\beta_j}(t) =&\ \mathsf{E}_{\boldsymbol{Z}}\, Z_j^2 \pi_0^\star(t)(1-\pi_0^\star(t))\\ \mathsf{Cov}(U_{i,\beta_j}(t), U_{i,\beta_{j'}}(t)) =&\ \mathsf{E}_{\boldsymbol{Z}}\, Z_j Z_{j'} \pi_0^\star(t)(1-\pi_0^\star(t)),\end{aligned}$$

hence the variances at 0 and τ are finite since we assume bounded covariates. Finally, the paths of $\boldsymbol{U}_i(t)$ are functions of the form

$$Z_j^c[\delta_i(t) - \pi_0^\star(t)],\ c = 0, 1,\ j = 1, \ldots, p$$

and these have finite variation as long as $\alpha_0^\star(t)$ does.

The derivatives of $\boldsymbol{U}_i$ with respect to $\alpha(t)$ and $\boldsymbol{\beta}$ equal

$$\frac{\partial \boldsymbol{U}_i}{\partial \alpha(t)} = \begin{pmatrix} -\pi_i^\star(t)(1-\pi_i^\star(t) \\ -z_{i1}\pi_i^\star(t)(1-\pi_i^\star(t) \\ \vdots \\ -z_{ip}\pi_i^\star(t)(1-\pi_i^\star(t) \end{pmatrix}$$

$$\frac{\partial \boldsymbol{U}_i}{\partial \boldsymbol{\beta}} = \begin{pmatrix} -z_{i1}\pi_i^\star(t)(1-\pi_i^\star(t) & \ldots & -z_{ip}\pi_i^\star(t)(1-\pi_i^\star(t) \\ -z_{i1}^2\pi_i^\star(t)(1-\pi_i^\star(t) & \ldots & -z_{i1}z_{ip}\pi_i^\star(t)(1-\pi_i^\star(t) \\ \vdots & \ddots & \vdots \\ -z_{i1}z_{ip}\pi_i^\star(t)(1-\pi_i^\star(t) & \ldots & -z_{ip}^2\pi_i^\star(t)(1-\pi_i^\star(t) \end{pmatrix}$$

and the expectation $\mathsf{E} - \frac{\partial \boldsymbol{U}_i(t)}{\partial \boldsymbol{\theta}(t)}$ equals $\boldsymbol{J}(t)$ for each t. Due to bounded covariates the variances of $\frac{\partial \boldsymbol{U}_i(t)}{\partial \boldsymbol{\theta}(t)}$ at 0 and τ are finite. The paths of $\frac{\partial \boldsymbol{U}_i(t)}{\partial \boldsymbol{\theta}(t)}$ have finite variation if $\alpha_0^\star(t)$ does. The assumptions of Lemma 4.8 are satisfied and $\hat{\boldsymbol{\theta}}(t)$ is uniformly consistent. ■

Theorem 4.10: *Let the baseline log odds function $\alpha_0^\star(t)$ have finite variation. The normalised score process evaluated at the true case-control parameter values $\frac{1}{\sqrt{n}}\boldsymbol{U}(\alpha_0^\star(t), \boldsymbol{\beta}_0, t)$ converges weakly to a zero-mean Gaussian process in $\ell^\infty(0, \tau]$*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{U}_i(t, \boldsymbol{\theta}_0(t)) \xrightarrow{\mathcal{D}} \mathbb{W}_{\boldsymbol{U}}(t). \tag{4.22}$$

The finite-dimensional covariance structure of $\mathbb{W}_{\boldsymbol{U}}(t)$ is given by the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{U}}(t_1, \ldots, t_K)$, see Lemma 4.2, consistent estimators of individual covariances are elements of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}(t_1, \ldots, t_K)$, see Lemma 4.3.
The normalised process of logistic regression estimators converges weakly to a zero-mean Gaussian process in $\ell^\infty(0, \tau]$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t)) \xrightarrow{\mathcal{D}} \mathbb{W}_{\boldsymbol{\theta}}(t). \tag{4.23}$$

The covariance structure of $\mathbb{W}_{\boldsymbol{\theta}}(t)$ is given by $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(t_1, \ldots, t_K)$ in Theorem 4.4 and can be consistently estimated by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}(t_1, \ldots, t_K)$, see Theorem 4.6. In particular,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)) \xrightarrow{\mathcal{D}} \mathbb{W}_{\boldsymbol{\beta}}(t) \tag{4.24}$$

with covariance structure $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t_1, \ldots, t_K)$ estimated by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_1, \ldots, t_K)$.

Proof: Convergence of the score process $\boldsymbol{U}$ follows as a special case of Lemma 4.8, validity of its assumptions was verified in the proof of Theorem 4.9. Finite dimensional distributions of the limiting process $\mathbb{W}_{\boldsymbol{U}}(t)$ were calculated and estimated earlier in Lemma 4.2 and Lemma 4.3.

The proof of convergence of the estimator process is done similarly as in Theorem 4.4. From the score equation we have that

$$\mathbf{0} = \frac{1}{\sqrt{n}}\boldsymbol{U}(t, \hat{\boldsymbol{\theta}}(t)) = \frac{1}{\sqrt{n}}\boldsymbol{U}(t, \boldsymbol{\theta}_0(t)) - \boldsymbol{J}(t)\ \sqrt{n}(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t)) + R_n(t),$$

where the leading term in $R_n(t)$ is given by

$$\frac{1}{\sqrt{n}}(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t))\ \frac{\partial^2 \boldsymbol{U}(t, \boldsymbol{\theta}(t))}{\partial \boldsymbol{\theta}(t)}\bigg|_{\boldsymbol{\theta}_0(t)} (\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t)).$$

Since $\hat{\boldsymbol{\theta}}(t)$ is uniformly consistent, R_n is $o_P(1)$ uniformly in t and

$$\sqrt{n}(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_0(t)) = \frac{1}{\sqrt{n}}\boldsymbol{J}^{-1}(t)\boldsymbol{U}(t, \boldsymbol{\theta}_0(t)) + o_P(1),$$

implying convergence to a Gaussian process. The covariance structure of $\mathbb{W}_{\boldsymbol{\theta}}(t)$ was calculated and estimated earlier (Theorems 4.4 and 4.6), the result for $\hat{\boldsymbol{\beta}}(t)$ is a simple consequence. ■

Now we need to generalize the concept of weights for the individual components of the combined logistic estimator, the weights are now nonnegative functions $w_1(t), \ldots, w_p(t), \tau_0 \le t \le \tau$. The values of w_j at observed failure times $t_1, \ldots, t_K$ are then used to construct the combined logistic estimator.

Definition 4.1: *Define the combined logistic estimator $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}$ as*

$$\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}} = \int_{\tau_0}^{\tau} \boldsymbol{W}(t)\hat{\boldsymbol{\beta}}(t)d\bar{N}(t), \tag{4.25}$$

where

$$\boldsymbol{W}(t) = \operatorname{diag}\left\{\frac{w_1(t)}{\int_{\tau_0}^{\tau} w_1(t)d\bar{N}(t)}, \ldots, \frac{w_p(t)}{\int_{\tau_0}^{\tau} w_p(t)d\bar{N}(t)}\right\}, \tag{4.26}$$

$w_j(t) \geq 0$ for all $t \geq \tau_0$ and the counting process $\bar{N}(t) = \sum_{i=1}^{n} \delta_i \mathbb{I}_{[T \leq t]}$.

Definition 4.1 is a direct generalization of the previous definition of $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}$ for fixed time points.

For proving the main result we will need one more technical lemma from Kulich & Lin (2000).

Lemma 4.11: *Let $A_n(t), A_n^\star(t)$ and $B_n(t)$ be three sequences of bounded processes on $[0, \tau]$. Suppose that*

a) $B_n(t)$ converges weakly to a tight limit $B(t)$ with almost surely continuous sample paths

b) $A_n(t)$ and $A_n^\star(t)$ are monotone in t

c) there exist processes $A(t)$ and $A^\star(t)$, both right continuous at 0 and left continuous at τ, such that $\sup_{0 \leq t \leq \tau} |A_n(t) - A(t)| \to_P 0$ and also $\sup_{0 \leq t \leq \tau} |A_n^\star(t) - A^\star(t)| \to_P 0$.

Then

$$\sup_{0 \leq t \leq \tau} \left| \int_0^t \{A_n(s) A_n^\star(s) - A(s) A^\star(s)\} \, dB_n(s) \right| \xrightarrow{P} 0.$$

Proof: The proof is given in Kulich & Lin (2000). ■

Theorem 4.12: *The combined logistic estimator (4.25) is consistent and asymptotically normal*

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}^\star_{\boldsymbol{\beta},\boldsymbol{W}}). \tag{4.27}$$

The asymptotic covariance matrix $\boldsymbol{\Sigma}^\star_{\boldsymbol{\beta},\boldsymbol{W}}$ consists of elements

$$\mathsf{C}(\tilde{\beta}_j, \tilde{\beta}_{j'}) = \int_{\tau_0}^{\tau} \int_{\tau_0}^{\tau} \mathsf{C}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right) d\mu_{W_j}(t) d\mu_{W_{j'}}(u), \ j, j' = 1, \ldots, p, \tag{4.28}$$

where $\mathsf{C}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ is the asymptotic covariance of components of $\hat{\boldsymbol{\beta}}(t)$ and $\hat{\boldsymbol{\beta}}(u)$ and $\mu_{W_j}(t) = \int_{\tau_0}^{t} w_j(s) \, d \,\mathsf{E}\, N_i(s) / \int_{\tau_0}^{\tau} w_j(s) \, d \,\mathsf{E}\, N_i(s)$ is the asymptotic cumulative weight of the j^{th} component till time t.

Note that the asymptotic covariances $\mathsf{C}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ are for any pair of times (t, u) elements of the asymptotic covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(t, u)$, which was defined in Theorem 4.4 more generally for K analysis times.

Proof: To simplify the notation denote for any vector $\boldsymbol{a}_{m\times 1}$ the vector of absolute values $(|a_1|,\dots,|a_m|)'$ by $|\boldsymbol{a}|$. Since $\int_{\tau_0}^{\tau} \boldsymbol{W}(t) d\bar{N}(t) = \boldsymbol{I}_{p\times p}$ and $d\bar{N}(t)$ is nondecreasing, we can write

$$\begin{aligned}
|\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}} - \boldsymbol{\beta}_0| &= \left| \int_{\tau_0}^{\tau} \boldsymbol{W}(t)\hat{\boldsymbol{\beta}}(t)\, d\bar{N}(t) - \int_{\tau_0}^{\tau} \boldsymbol{W}(t)\boldsymbol{\beta}_0\, d\bar{N}(t) \right| \\
&= \left| \int_{\tau_0}^{\tau} \boldsymbol{W}(t)(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0)\, d\bar{N}(t) \right| \\
&\le \int_{\tau_0}^{\tau} \boldsymbol{W}(t) \left| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0 \right| d\bar{N}(t) \\
&\le \int_{\tau_0}^{\tau} \boldsymbol{W}(t) \sup_{\tau_0 \le t \le \tau} \left| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0 \right| d\bar{N}(t) \\
&= \left[\int_{\tau_0}^{\tau} \boldsymbol{W}(t)\, d\bar{N}(t) \right] \sup_{\tau_0 \le t \le \tau} \left| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0 \right| \\
&= \sup_{\tau_0 \le t \le \tau} \left| \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0 \right|.
\end{aligned}$$

Uniform consistency of $\hat{\boldsymbol{\beta}}(t)$ implies consistency of $\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}$.

To avoid cumbersome matrix notation we shall prove the asymptotic normality only for $p = 1$ in detail. The generalization to $p > 1$ is straightforward. We have

$$\begin{aligned}
\sqrt{n}(\tilde{\beta}_W - \beta_0) &= \int_{\tau_0}^{\tau} W(t)\sqrt{n}\left(\hat{\beta}(t) - \beta_0\right) d\bar{N}(t) \\
&= \int_{\tau_0}^{\tau} \sqrt{n}\left(\hat{\beta}(t) - \beta_0\right) dN_W(t),
\end{aligned}$$

where $N_W(t) = \int_{\tau_0}^{t} W(s)\, d\bar{N}(s)$ is the cumulative weight process. This process can be represented as

$$N_W(t) = \frac{\frac{1}{n}\sum_{i=1}^{n} \int_{\tau_0}^{t} w(s)\, dN_i(s)}{\frac{1}{n}\sum_{i=1}^{n} \int_{\tau_0}^{\tau} w(s)\, dN_i(s)}, \tag{4.29}$$

the numerator and denominator converge in probability to their respective expectations uniformly in t by Lemma 4.8 and therefore $N_W(t)$ converges in probability to some nonrandom function $\mu_W(t)$ uniformly in t.

Denote further $\sqrt{n}(\hat{\beta}(t) - \beta_0)$ by $B_n(t)$; we know that $B_n(t)$ converges weakly to a zero-mean Gaussian process $\mathbb{W}_\beta$. Using integration by parts

$$
\begin{aligned}
\sqrt{n}(\tilde{\beta}_W - \beta_0) &= \int_{\tau_0}^{\tau} B_n(t) dN_W(t) \\
&= B_n(\tau)N_W(\tau) - B_n(\tau_0)N_W(\tau_0) - \int_{\tau_0}^{\tau} N_W(t-)\, dB_n(t) \\
&= B_n(\tau)N_W(\tau) - B_n(\tau_0)N_W(\tau_0) - \int_{\tau_0}^{\tau} \mu_W(t-)\, dB_n(t) + o_P(1),
\end{aligned}
$$

since $\left| \int_{\tau_0}^{\tau} N_W(t)\, dB_n(t) - \int_{\tau_0}^{\tau} \mu_W(t)\, dB_n(t) \right|$ tends to 0 in probability by Lemma 4.11. Moreover, we have convergence of $\frac{1}{n}\bar{N}(s)$ to the continuous distribution function of failure time T (censoring only occurs at the end of the study), therefore both the numerator and the denominator of (4.29) have continuous limits. The denominator is bounded away from zero so μ_W is continuous and thus $\mu_W(t-) = \mu_W(t)$.

Applying integration by parts again gives

$$
\int_{\tau_0}^{\tau} \mu_W(t)\, dB_n(t) = B_n(\tau)\mu_W(\tau) - B_n(\tau_0)\mu_W(\tau_0) - \int_{\tau_0}^{\tau} B_n(t-)\, d\mu_W(t)
$$

and therefore

$$
\begin{aligned}
\sqrt{n}(\tilde{\beta}_W - \beta_0) &= B_n(\tau)\left[N_W(\tau) - \mu_W(\tau)\right] - B_n(\tau_0)\left[N_W(\tau_0) - \mu_W(\tau_0)\right] \\
&\quad + \int_{\tau_0}^{\tau} B_n(t-)\, d\mu_W(t) + o_P(1) \\
&= \int_{\tau_0}^{\tau} B_n(t-)\, d\mu_W(t) + o_P(1) \\
&= \int_{\tau_0}^{\tau} \mathbb{W}_\beta(t)\, d\mu_W(t) + o_P(1),
\end{aligned} \tag{4.30}
$$

since $[N_W(\tau) - \mu_W(\tau)]$ and $[N_W(\tau_0) - \mu_W(\tau_0)]$ both converge to 0 in probability (Lemma 4.8) and $\left| \int_{\tau_0}^{\tau} B_n(t-)\, d\mu_W(t) - \int_{\tau_0}^{\tau} \mathbb{W}_\beta(t)\, d\mu_W(t) \right|$ tends to 0 almost surely (a result taken from the proof Lemma 4.11, see Kulich & Lin (2000)). The asymptotic normality is now clear because the last integral in (4.30) can be expressed as a sum of jointly normally distributed random variables

$$
\int_{\tau_0}^{\tau} \mathbb{W}_\beta(t)\, d\mu_W(t) = \mathbb{W}_\beta(\tau)\mu_W(\tau) - \mathbb{W}_\beta(\tau_0)\mu_W(\tau_0) - \int_{\tau_0}^{\tau} \mu_W(t)\, d\mathbb{W}_\beta.
$$

For $p > 1$ the proof proceeds in the same way as above, the process $\mathbb{W}_{\boldsymbol{\beta}}(t)$ still has mean zero and its covariance structure is given by

$$\mathsf{Cov}(\mathbb{W}_{\beta_j}, \mathbb{W}_{\beta_{j'}}) = \mathsf{C}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right),$$

see Theorem 4.4. Therefore the elements of the asymptotic variance matrix of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}} - \boldsymbol{\beta}_0)$ are given by

$$\begin{aligned}\mathsf{C}(\tilde{\beta}_j, \tilde{\beta}_{j'}) &= \mathsf{Cov}\left(\int_{\tau_0}^{\tau} \mathbb{W}_{\beta_j}(t) d\mu_{W_j}(t), \int_{\tau_0}^{\tau} \mathbb{W}_{\beta_{j'}}(u) d\mu_{W_{j'}}(u)\right) \\ &= \mathsf{E}\left(\int_{\tau_0}^{\tau} \mathbb{W}_{\beta_j}(t) d\mu_{W_j}(t) \int_{\tau_0}^{\tau} \mathbb{W}_{\beta_{j'}}(u) d\mu_{W_{j'}}(u)\right) \\ &= \int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau} \mathsf{Cov}\left(\mathbb{W}_{\beta_j}(t), \mathbb{W}_{\beta_{j'}}(u)\right) d\mu_{W_j}(t) d\mu_{W_{j'}}(u).\end{aligned}$$

■

Theorem 4.13: *The asymptotic variance matrix of $\tilde{\boldsymbol{\beta}}$ can be consistently estimated by a $p \times p$ matrix $\hat{\boldsymbol{\Sigma}}^{\star}_{\boldsymbol{\beta},\boldsymbol{W}}$ with elements*

$$\hat{\mathsf{C}}(\tilde{\beta}_j, \tilde{\beta}_{j'}) = \int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau} W_j(t) \hat{\mathsf{C}}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right) W_{j'}(u) d\bar{N}(t) d\bar{N}(u), \tag{4.31}$$

where $\hat{\mathsf{C}}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ are estimated asymptotic covariances of components of $\hat{\boldsymbol{\beta}}(t)$ and $\hat{\boldsymbol{\beta}}(u)$, that is elements of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(t_1, \ldots, t_K)$.

Note that formula (4.31) can be rewritten as

$$\hat{\mathsf{C}}(\tilde{\beta}_j, \tilde{\beta}_{j'}) = \sum_{k=1}^{K}\sum_{k'=1}^{K} W_j(t_k) \hat{\mathsf{C}}\left(\hat{\beta}_j(t_k), \hat{\beta}_{j'}(t_{k'})\right) W_{j'}(t_{k'}),$$

which is the estimator of the elements of $\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{W}(t_1,\ldots,t_K)}$ (4.18) calculated at the observed failure times.

Proof: By Theorem 4.6 and Lemma 4.3, $\hat{\mathsf{C}}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ is a consistent estimator of $\mathsf{C}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ for any fixed times t and u. The proof utilises consistency of $\hat{\boldsymbol{\beta}}(t)$ and Taylor expansion of $\hat{\mathsf{C}}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ around true parameter values. By Theorem 4.9, the estimator $\hat{\boldsymbol{\beta}}(t)$ is uniformly consistent in t, therefore $\hat{\mathsf{C}}(\hat{\beta}_j(t), \hat{\beta}_{j'}(u))$ is also uniformly consistent. Thus

$$
\begin{aligned}
\hat{\mathsf{C}}(\tilde{\beta}_j, \tilde{\beta}_{j'}) &= \int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau} W_j(t)\hat{\mathsf{C}}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right) W_{j'}(u) d\bar{N}(t) d\bar{N}(u) \\
&= \int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau} \hat{\mathsf{C}}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right) dN_{W_j}(t) dN_{W_{j'}}(u) \\
&= \int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau} \left[\mathsf{C}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right) + R_n(t,u)\right] dN_{W_j}(t) dN_{W_{j'}}(u) \\
&= \int_{\tau_0}^{\tau}\int_{\tau_0}^{\tau} \mathsf{C}\left(\hat{\beta}_j(t), \hat{\beta}_{j'}(u)\right) d\mu_{W_j}(t) d\mu_{W_{j'}}(u) + o_P(1),
\end{aligned}
$$

since $R_n(t,u)$ is $o_P(1)$ uniformly in t and u and $N_{W_j}(t)$ converges in probability to $\mu_{W_j}(t)$ uniformly in t. ■

True weights (weighting processes) $\boldsymbol{W}(t)$, which were used in all previous statements since Theorem 4.9, depend through asymptotic variance matrices on unknown parameters. We need to show that true weights can be replaced with estimated weights when constructing the combined logistic estimator without affecting the results.

Theorem 4.14: *Let the optimal weights $\boldsymbol{W}^{\mathrm{opt}}(t)$ depend on the covariance function of $\hat{\boldsymbol{\beta}}(t)$ through some continuous function ϕ, i.e. $\boldsymbol{W}^{\mathrm{opt}}(t) = \phi(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(s,t))$. Define the estimated weights as $\hat{\boldsymbol{W}}^{\mathrm{opt}}(t) = \phi(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(s,t))$. Then*

$$
\sqrt{n}\left(\tilde{\boldsymbol{\beta}}^{\mathrm{opt}}_{\hat{\boldsymbol{W}}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{D}} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}^{\star}_{\boldsymbol{\beta},\boldsymbol{W}^{\mathrm{opt}}}), \tag{4.32}
$$

where $\tilde{\boldsymbol{\beta}}^{\mathrm{opt}}_{\hat{\boldsymbol{W}}} = \int_{\tau_0}^{\tau} \hat{\boldsymbol{W}}^{\mathrm{opt}}(t)\hat{\boldsymbol{\beta}}(t)d\bar{N}(t)$ is the optimal combined logistic estimator computed using the estimated weights.

Proof: Similarly as in the proof of Theorem 4.7 we can write

$$
\begin{aligned}
\sqrt{n}(\tilde{\boldsymbol{\beta}}^{\mathrm{opt}}_{\hat{\boldsymbol{W}}} - \boldsymbol{\beta}_0) &= \sqrt{n}\left(\int_{\tau_0}^{\tau} \hat{\boldsymbol{W}}^{\mathrm{opt}}(t)\hat{\boldsymbol{\beta}}(t)\, d\bar{N}(t) - \boldsymbol{\beta}_0\right) \\
&= \int_{\tau_0}^{\tau} \hat{\boldsymbol{W}}^{\mathrm{opt}}(t)\sqrt{n}\left(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\right) d\bar{N}(t) \\
&= \int_{\tau_0}^{\tau} \boldsymbol{W}^{\mathrm{opt}}(t)\sqrt{n}\left(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\right) d\bar{N}(t) \\
&\quad + \int_{\tau_0}^{\tau} \left(\hat{\boldsymbol{W}}^{\mathrm{opt}}(t) - \boldsymbol{W}^{\mathrm{opt}}(t)\right)\sqrt{n}\left(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\right) d\bar{N}(t)
\end{aligned}
$$

$$
\begin{aligned}
&= \sqrt{n}\left(\int_{\tau_0}^{\tau} \boldsymbol{W}^{\mathrm{opt}}(t)\hat{\boldsymbol{\beta}}(t)\, d\bar{N}(t) - \boldsymbol{\beta}_0\right) \\
&\quad + \int_{\tau_0}^{\tau} \Big(\boldsymbol{\phi}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(s,t)) - \boldsymbol{\phi}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(s,t))\Big) \sqrt{n}\left(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\right) d\bar{N}(t) \\
&= \sqrt{n}\left(\tilde{\boldsymbol{\beta}}_{\boldsymbol{W}}^{\mathrm{opt}} - \boldsymbol{\beta}_0\right) + o_P(1),
\end{aligned}
$$

since $\bar{N}(t)$ is nondecreasing, $\boldsymbol{\phi}$ is continuous, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(s,t)$ and $\hat{\boldsymbol{\beta}}(t)$ are uniformly consistent estimators of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(s,t)$ and $\boldsymbol{\beta}_0$, respectively, and $\sqrt{n}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0)$ converges weakly to a Gaussian process. Therefore

$$
\begin{aligned}
&\int_{\tau_0}^{\tau} \Big(\boldsymbol{\phi}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(s,t)) - \boldsymbol{\phi}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(s,t))\Big) \sqrt{n}\left(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\right) d\bar{N}(t) \\
&\qquad \leq \int_{\tau_0}^{\tau} \sup_t \Big|\boldsymbol{\phi}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(s,t)) - \boldsymbol{\phi}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(s,t))\Big| \sup_t \Big|\sqrt{n}(\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0)\Big|\, d\bar{N}(t) \\
&\qquad = o_P(1).
\end{aligned}
$$

■

Similarly as for fixed times of analyses it follows that the elements of the asymptotic variance matrix $\boldsymbol{\Sigma}^{\star}_{\boldsymbol{\beta},\boldsymbol{W}}$ can be consistently estimated using the estimated weights by $\boldsymbol{\Sigma}^{\star}_{\boldsymbol{\beta},\hat{\boldsymbol{W}}}$. The elements of $\boldsymbol{\Sigma}^{\star}_{\boldsymbol{\beta},\hat{\boldsymbol{W}}}$ are given by formulas (4.20) calculated in the observed failure times.

By proving asymptotic normality of the combined logistic estimator we have established the basis for building asymptotic tests and confidence intervals for the regression parameter $\boldsymbol{\beta}$. The section on theoretical results is completed.

Chapter 5

A numerical study

In a simulation study in Chapter 3 we have shown that small values of the subcohort sampling probability α can influence the performance of standard case-cohort estimators. To support the idea of the combined logistic estimator (CLE) we have conducted another simulation study, where we compared our estimator with traditional approaches. The simulation design, comparison of estimators and performance reports are provided in this chapter.

5.1 Simulation design

We performed $2 \times 2 \times 2 = 8$ sets of simulation studies combining data from the proportional hazards model and from the proportional odds model, constant and uniform censoring and small and large cohorts. For each of the simulation designs we generated 1000 full cohorts, selected the subcohort by independent Bernoulli sampling and computed mean estimate, mean standard error (based on the asymptotic distribution) and estimated (empirical) standard error of the estimate, mean 95% confidence interval coverage, bias and mean squared error for all the compared estimators.

There were always two dependent covariates, a binary covariate Z_1 with $\mathrm{P}(Z_1 = 1) = 0.35$ and a truncated normal covariate $Z_2 \sim \mathrm{N}(\mu_{Z_2}, \sigma^2_{Z_2})$ with $\mu_{Z_2} = 1.5Z_1$ and $\sigma^2_{Z_2} = (1 + 1.5Z_1)^2$, truncated at $\mu_{Z_2} \pm 3\sigma_{Z_2}$. Truncation was introduced in order to satisfy the condition of bounded covariates, which is imposed by some of the estimating techniques. The true regression parameters were set to $\boldsymbol{\beta}_0 = (2.3, 1.2)'$.

Censoring mechanisms were constant censoring and independent uniform censoring. We always kept the expected number of cases to be 100 and the expected number of controls sampled to the case-cohort study to be also 100. The small cohort consisted of 10 000 and the large cohort of 300 000 individuals. The subcohort sampling rates were therefore 0.01 for the small cohort and 0.0003 for the large cohort.

In simulations for the large cohort we only estimated model parameters based on case-cohort data. In simulations for the small cohort we estimated model parameters based on the case-cohort data as well as based on the full cohort. Using case-cohort data, we compared the combined logistic estimator to the original Prentice estimator, the estimator proposed by Lu & Tsiatis (2006) and to the case-control estimator obtained from a logistic regression model at the end of the study. When full data were analysed, we used the Cox partial likelihood estimator and the estimator proposed by Chen et al. (2002) as representants of traditional estimators, since the Prentice estimator and the estimator by Lu & Tsiatis are their case-cohort variants. The threshold parameter of the combined logistic estimator k_0 was set to 30 for all simulations, that means each combined logistic estimator was based on 70 logistic regression models.

5.2 Results

Constant censoring

In the first set of simulations we generated data from the proportional odds model and applied censoring by a constant value. In 1000 repetitions we observed on average 99.46 cases and 100.96 sampled subcohort subjects (99.95 subcohort controls) out of 10 000 observations in the full cohort. Table 5.1 shows performance of the estimators for case-cohort and full data.

We can see the same pattern as in Chapter 3. All estimators perform reasonably well with full data, even the Cox partial likelihood estimator – hazard ratios estimated by the Cox model closely approximate odds ratios specified by the proportional odds model. The estimators of Prentice and Lu & Tsiatis encounter problems with case-cohort data. There is a clear bias (as high as 10% and 16%) and low confidence interval coverage. Moreover, model-based standard errors seem to be underestimated compared to their empirical counterparts. Note that the sampling probability and the probability of an event are still relatively large here (0.01).

The combined logistic estimator behaves well with full data and is the best estimator (measured by mean squared errors) with case-cohort data. It is only minimally biased, the model-based and sample standard errors agree and confidence intervals cover the true parameters better than with traditional estimators. The case-control estimator from the end of the study also performs quite well and in many aspects it is comparable to the combined logistic estimator. This may be due to special censoring patters, however more detailed studies would be necessary to support this hypothesis.

Table 5.1: Simulation summary I – Proportional odds model, constant censoring, cohort size 10 000, sampling probability $\alpha = 0.01$.

(a) Case-cohort data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.049	2.349	0.384	0.403	95.300	0.165
	β_2	0.100	1.300	0.345	0.456	88.200	0.218
Lu	β_1	0.124	2.424	0.400	0.433	95.200	0.203
	β_2	0.194	1.394	0.372	0.508	86.100	0.296
CLE	β_1	0.005	2.305	0.360	0.371	94.483	0.137
	β_2	-0.001	1.199	0.279	0.311	92.477	0.096
Case – Control	β_1	0.063	2.379	0.373	0.376	95.386	0.145
	β_2	0.046	1.246	0.300	0.314	95.186	0.100

(b) Full data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Cox	β_1	0.006	2.306	0.272	0.287	95.300	0.083
	β_2	-0.024	1.176	0.151	0.156	93.700	0.025
Chen	β_1	0.026	2.326	0.272	0.289	95.000	0.084
	β_2	0.000	1.200	0.156	0.162	94.000	0.026
CLE	β_1	0.022	2.322	0.273	0.288	94.960	0.083
	β_2	-0.004	1.196	0.157	0.161	94.355	0.026
Case – Control	β_1	0.025	2.325	0.273	0.289	95.161	0.084
	β_2	0.001	1.201	0.157	0.162	94.254	0.026

For the second set of simulations we increased the cohort size to 300 000 and in 1000 repetitions we observed on average 100.76 cases and 100.67 sampled subcohort subjects (100.64 subcohort controls). Results are summarised in Table 5.2. Problems of traditional estimators are more pronounced when the sampling probability drops to 0.0003, while the combined logistic estimator still shows small bias and MSE. The Prentice estimator designed for the proportional hazards model is surprisingly better than the estimator by Lu & Tsiatis, which is much more general.

Table 5.2: Simulation summary II – Proportional odds model, constant censoring, cohort size 300 000, sampling probability $\alpha = 0.0003$.

Case-cohort data								
Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.049	2.349	0.385	0.435	92.900	0.192
	β_2	0.110	1.310	0.347	0.461	88.000	0.224
Lu	β_1	0.112	2.416	0.410	0.495	92.800	0.258
	β_2	0.228	1.428	0.404	0.614	88.800	0.429
CLE	β_1	−0.001	2.299	0.359	0.387	93.173	0.150
	β_2	−0.003	1.197	0.276	0.300	93.876	0.090
Case – Control	β_1	0.050	2.350	0.372	0.391	94.277	0.156
	β_2	0.035	1.235	0.296	0.299	95.582	0.091

We performed the same pair of studies with data following the proportional hazards model to see how the combined logistic estimator would perform in this situation. In 1000 repetitions we observed on average 96.79 cases and 100.88 sampled subcohort subjects (99.98 subcohort controls) out of 10 000 observations in the full cohort. In the last study for constant censoring, with 300 000 individuals, we observed on average 97.83 cases and 101.43 sampled subcohort subjects (101.40 subcohort controls).

Table 5.3: Simulation summary III – Proportional hazards model, constant censoring, cohort size 10 000, sampling probability $\alpha = 0.01$.

(a) Case-cohort data								
Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.059	2.359	0.388	0.426	93.700	0.185
	β_2	0.111	1.311	0.344	0.477	85.600	0.240
Lu	β_1	0.127	2.427	0.402	0.453	93.000	0.221
	β_2	0.196	1.396	0.370	0.521	84.900	0.310
CLE	β_1	0.017	2.317	0.364	0.396	92.871	0.157
	β_2	−0.004	1.204	0.280	0.307	93.574	0.094
Case – Control	β_1	0.070	2.370	0.377	0.400	93.474	0.165
	β_2	0.049	1.249	0.301	0.301	95.482	0.098

(b) Full data							
Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Cox	β_1	0.020	2.320	0.298	0.298	94.400	0.088
	β_2	-0.003	1.197	0.155	0.155	95.300	0.022
Chen	β_1	0.040	2.340	0.278	0.298	94.600	0.090
	β_2	0.022	1.222	0.158	0.155	94.300	0.025
CLE	β_1	0.033	2.333	0.279	0.293	95.147	0.087
	β_2	0.017	1.217	0.159	0.153	95.046	0.024
Case – Control	β_1	0.036	2.336	0.279	0.294	95.046	0.087
	β_2	0.023	1.223	0.160	0.155	94.641	0.025

Table 5.4: Simulation summary IV – Proportional hazards model, constant censoring, cohort size 300 000, sampling probability $\alpha = 0.0003$.

Case-cohort data							
Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.054	2.354	0.388	0.406	94.000	0.167
	β_2	0.100	1.300	0.343	0.456	87.500	0.217
Lu	β_1	0.121	2.421	0.409	0.462	93.500	0.228
	β_2	0.216	1.416	0.397	0.608	88.800	0.417
CLE	β_1	-0.001	2.299	0.362	0.368	94.052	0.135
	β_2	-0.007	1.193	0.277	0.303	92.440	0.095
Case – Control	β_1	0.051	2.351	0.374	0.371	95.161	0.140
	β_2	0.033	1.233	0.297	0.306	94.355	0.095

Tables 5.3 and 5.4 show performance of the estimators for case-cohort and full data. We can see that the combined logistic estimator performs well already with the smaller cohort size 10 000, that means event probability 0.01. The odds ratios closely approximate relative risks and the estimator shows better performance than the original Prentice's estimator for the proportional hazards model. The case-control estimator also shows very good results and finally all estimators behave well with full data.

Independent uniform censoring

The remaining four sets of simulations were carried out in the same way as before but censoring occurred randomly – censoring time was simulated as a uniformly distributed random variable, independent on survival time. Let us first report the results with data from the proportional odds model with 10 000 subjects in the full cohort. In 1000 repetitions we observed on average 97.55 cases and 101.09 sampled subcohort subjects (100.08 subcohort controls), for a detailed report see Table 5.5.

All estimators perform reasonably well with full data, the problems of estimators by Prentice and Lu & Tsiatis with case-cohort data remain roughly the same as with constant censoring. We can see high bias and wrong standard errors and interval coverage, especially for the second parameter β_2 belonging to the truncated normal covariate.

The combined logistic estimator still behaves well with full as well as with case-cohort data. There is minimal bias, good agreement in model-based and empirical standard errors and good coverage probability for confidence intervals. Surprisingly, the case-control logistic regression estimator outperforms both traditional case-cohort estimators even when random censoring is present.

Table 5.5: Simulation summary V – Proportional odds model, uniform censoring, cohort size 10 000, sampling probability $\alpha = 0.01$.

(a) Case-cohort data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.015	2.315	0.402	0.424	93.200	0.180
	β_2	0.109	1.309	0.349	0.468	86.000	0.230
Lu	β_1	0.115	2.415	0.427	0.470	92.000	0.233
	β_2	0.225	1.425	0.384	0.544	83.300	0.346
CLE	β_1	−0.025	2.275	0.359	0.369	94.472	0.137
	β_2	0.004	1.204	0.279	0.299	92.171	0.089
Case – Control	β_1	0.026	2.326	0.372	0.371	95.779	0.138
	β_2	0.045	1.245	0.300	0.299	95.779	0.091

(b) Full data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Cox	β_1	-0.022	2.288	0.272	0.275	94.600	0.076
	β_2	-0.012	1.178	0.154	0.150	95.600	0.023
Chen	β_1	0.013	2.313	0.272	0.276	94.800	0.076
	β_2	0.009	1.209	0.160	0.157	95.400	0.025
CLE	β_1	-0.005	2.295	0.273	0.274	94.726	0.075
	β_2	-0.009	1.191	0.159	0.154	95.842	0.024
Case – Control	β_1	-0.003	2.297	0.273	0.274	94.726	0.075
	β_2	-0.005	1.195	0.159	0.155	95.740	0.024

With cohort size equal 300 000 we observed on average 105.17 cases and 100.36 sampled subcohort subjects (100.31 subcohort controls). Results summarised in Table 5.6 confirm what we have already seen before – the combined logistic estimator performs clearly better than Prentice or Lu & Tsiatis estimators.

Also here the Prentice estimator designed for the proportional hazards model is better than the estimator by Lu & Tsiatis. The case-control estimator remains a good alternative showing only slightly worse results than the combined logistic estimator.

Table 5.6: Simulation summary VI – Proportional odds model, uniform censoring, cohort size 300 000, sampling probability $\alpha = 0.0003$.

Case-cohort data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.059	2.359	0.400	0.448	92.500	0.204
	β_2	0.105	1.305	0.347	0.464	86.100	0.226
Lu	β_1	0.160	2.460	0.435	0.538	91.900	0.315
	β_2	0.265	1.465	0.420	0.670	85.200	0.519
CLE	β_1	0.007	2.307	0.355	0.378	93.970	0.143
	β_2	0.002	1.202	0.272	0.301	92.764	0.090
Case – Control	β_1	0.060	2.360	0.368	0.384	94.271	0.151
	β_2	0.038	1.238	0.293	0.302	94.573	0.092

Finally, the last two sets of simulations involve data from the proportional hazards model. For the small cohort of 10 000 individuals we observed on average 102.33 cases and 101.48 sampled subcohort subjects (100.45 subcohort controls), for the larger cohort size it was 102.02 cases and 101.31 sampled subcohort subjects (101.27 subcohort controls).

Table 5.7: Simulation summary VII – Proportional hazards model, uniform censoring, cohort size 10 000, sampling probability $\alpha = 0.01$.

(a) Case-cohort data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.049	2.349	0.399	0.436	93.000	0.192
	β_2	0.105	1.305	0.347	0.448	86.900	0.211
Lu	β_1	0.144	2.444	0.424	0.482	92.900	0.252
	β_2	0.213	1.413	0.379	0.518	84.000	0.313
CLE	β_1	−0.002	2.298	0.355	0.379	93.921	0.144
	β_2	0.002	1.202	0.276	0.290	94.428	0.084
Case – Control	β_1	0.056	2.356	0.369	0.381	94.630	0.148
	β_2	0.046	1.246	0.298	0.295	96.454	0.089

(b) Full data

Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Cox	β_1	0.017	2.317	0.268	0.273	95.100	0.075
	β_2	0.003	1.203	0.150	0.149	94.600	0.022
Chen	β_1	0.045	2.345	0.269	0.275	96.000	0.078
	β_2	0.037	1.237	0.157	0.158	93.000	0.026
CLE	β_1	0.022	2.322	0.269	0.268	95.799	0.072
	β_2	0.014	1.214	0.156	0.154	94.980	0.024
Case – Control	β_1	0.025	2.325	0.269	0.268	96.004	0.072
	β_2	0.019	1.219	0.156	0.155	94.262	0.024

Table 5.8: Simulation summary VIII – Proportional hazards model, uniform censoring, cohort size 300 000, sampling probability $\alpha = 0.0003$.

Case-cohort data							
Method	Par.	Bias	Mean Est.	Average st. err.	Empirical st. err.	95% CI coverage	MSE
Prentice	β_1	0.050	2.350	0.403	0.446	91.900	0.202
	β_2	0.137	1.337	0.348	0.474	85.300	0.243
Lu	β_1	0.155	2.455	0.438	0.549	90.700	0.325
	β_2	0.308	1.508	0.422	0.681	84.300	0.558
CLE	β_1	-0.011	2.289	0.356	0.362	94.657	0.131
	β_2	0.012	1.212	0.275	0.309	92.742	0.095
Case – Control	β_1	0.039	2.339	0.369	0.362	96.069	0.132
	β_2	0.050	1.250	0.295	0.309	95.363	0.098

The results reported in Tables 5.7 and 5.8 again confirm what we have seen with constant censoring. All estimators are good with full data. The combined logistic estimator outperforms both traditional estimators on case-cohort data, the approximation of hazard ratios with odds ratios works well.

Summary

Simulation results show that the combined logistic estimator is a useful alternative to classical approaches when analysing case-cohort data following the proportional hazards or the proportional odds model. Compared to traditional estimators, the combined logistic estimator is less biased and has lower MSE on case-cohort data, its performance on full cohort data does not substantially differ from the other estimators.

Theoretically the combined logistic estimator should show better performance than the case-control logistic regression estimator from the end of the study, since the combined logistic estimator has the weights optimised for variance. In practice we have seen that the case-control estimator is only slightly worse when we compare mean squared errors. This makes the case-control estimator an interesting option in situations where computational time is an issue. Further research would be necessary to explain this behaviour in more detail.

Chapter 6

Summary and discussion

In the thesis we were dealing with regression models and parameter estimation in case-cohort studies. After a brief introduction to regression models in survival analysis and to the case-cohort design we described in detail how standard estimation techniques from survival analysis are usually adapted to case-cohort data. We reviewed the principle of weighting estimating equations with inverse sampling probabilities and theoretical properties of case-cohort estimators. We also compared parameter estimation in the proportional odds model with parameter estimation in logistic regression and found a close relationship.

We were particularly interested in situations where only a small fraction of individuals is sampled for the analysis from a large cohort. The case-cohort design is most useful here as it can save most of the costs. In a simulation study in Chapter 3 we performed a sensitivity analysis on α, the probability of subcohort sampling, and demonstrated problems of current case-cohort estimators when dealing with small sampling probabilities. Motivated by the fact that logistic regression can be applied to case-control data without inverse probability weighting we developed a new estimator of parameters in the proportional odds model. The estimator is based on fitting a logistic regression model repeatedly after each failure (event) and using a convex linear combination of the results as the combined logistic estimator. This way we retain time information that cannot be captured in any single case-control logistic analysis.

In Chapter 4 we showed consistency and asymptotic normality of the combined logistic estimator and in Chapter 5 we studied its performance in comparison with classical case-cohort estimators. We found out that the combined logistic estimator performed better than traditional estimators in all the settings we considered. Most notably the combined logistic estimator was always only slightly biased, had better confidence interval coverage and smaller MSE than traditional estimators. Since we were focused on situations with low event rates 0.01 and 0.0003, the relative risk was well approximated by the odds ratio and the combined logistic estimator also performed well for data generated from the proportional hazards model.

There are several open problems and topics for further work. The whole procedure was developed under the assumptions of constant censoring and fixed covariates. The procedure will still work with a general censoring distribution independent of the covariates. The only difference is due to subjects who would fail until τ but are censored earlier instead.

Consider an example with a single binary covariate for illustration. With censoring only at the end of study, the odds ratio can be expressed as

$$\frac{\mathrm{P}(Z=1,T<\tau)\mathrm{P}(Z=0,T>\tau)}{\mathrm{P}(Z=1,T>\tau)\mathrm{P}(Z=0,T<\tau)}=\frac{\mathrm{P}(T<\tau|Z=1)}{\mathrm{P}(T<\tau|Z=0)}\cdot\frac{1-\mathrm{P}(T<\tau|Z=0)}{1-\mathrm{P}(T<\tau|Z=1)}.$$

With a censoring variable C, this odd ratio becomes

$$\frac{\mathrm{P}(Z=1,T<\tau\wedge C)\mathrm{P}(Z=0,T>\tau\wedge C)}{\mathrm{P}(Z=1,T>\tau\wedge C)\mathrm{P}(Z=0,T<\tau\wedge C)},$$

where $\tau\wedge C$ denotes $\min(\tau,C)$. This formula can be expressed as

$$\frac{q_1}{q_0}\cdot\frac{\mathrm{P}(T<\tau|Z=1)}{\mathrm{P}(T<\tau|Z=0)}\cdot\frac{1-q_0\mathrm{P}(T<\tau|Z=0)}{1-q_1\mathrm{P}(T<\tau|Z=1)},$$

where $q_z=\mathrm{P}(T<C|T<\tau,Z=z)=\mathrm{P}(\delta=1|T<\tau,Z=z)$ for $z=0,1$. With small event probabilities, the last fraction equals approximately 1. The bias is then primarily induced by q_1/q_0 and can be expressed as

$$\frac{q_1}{q_0}=\frac{F(\tau|Z=0)}{F(\tau|Z=1)}\cdot\frac{\int_0^\infty F(u\wedge\tau|Z=1)g(u|Z=1)du}{\int_0^\infty F(u\wedge\tau|Z=0)g(u|Z=0)du},$$

where $F(\cdot)$ denotes the distribution function of T and $g(\cdot)$ the density of C. Since $F(\tau|Z=z)=\exp\{\alpha(\tau)+\beta z\}/(1+\exp\{\alpha(\tau)+\beta z\})$ in the proportional odds model, we have

$$\begin{aligned}\frac{q_1}{q_0}&=\frac{\frac{\exp\{\alpha(\tau)\}}{1+\exp\{\alpha(\tau)\}}}{\frac{\exp\{\alpha(\tau)+\beta z\}}{1+\exp\{\alpha(\tau)+\beta z\}}}\cdot\frac{\int_0^\infty\frac{\exp\{\alpha(u\wedge\tau)+\beta z\}}{1+\exp\{\alpha(u\wedge\tau)+\beta z\}}g(u|Z=1)du}{\int_0^\infty\frac{\exp\{\alpha(u\wedge\tau)\}}{1+\exp\{\alpha(u\wedge\tau)\}}g(u|Z=0)du}\\&=\exp\{-\beta z\}\frac{1+\exp\{\alpha(\tau)+\beta z\}}{1+\exp\{\alpha(\tau)\}}\cdot\frac{\exp\{\beta z\}\,\mathsf{E}\left[\frac{\exp\{\alpha(C\wedge\tau)\}}{1+\exp\{\alpha(C\wedge\tau)+\beta z\}}\,|Z=1\right]}{\mathsf{E}\left[\frac{\exp\{\alpha(C\wedge\tau)\}}{1+\exp\{\alpha(C\wedge\tau)\}}\,|Z=0\right]},\end{aligned}$$

where the expectation is taken with respect to C. Again, with low event probabilities we have approximately $\frac{1+\exp\{\alpha(\tau)+\beta z\}}{1+\exp\{\alpha(\tau)\}} \approx 1$. If $g(\cdot)$ and therefore the expectation do not depend on Z, the ratio q_1/q_0 is approximately 1.

If the censoring variable depends on the covariates or the covariates are time-dependent, the formulas for covariances between individual estimators need to take this into account. General at-risk processes could also be allowed, but the individual scores would then be based on different individuals every time. However, with corrected covariances between the estimators, the rest of the procedure remains the same.

In the development of the estimator we only considered diagonal weighting matrices for combining individual logistic estimators. We believe that efficiency gains from allowing more general structures for the weights would be minimal. Moreover, all elements of the weighting matrices need to be estimated from the data and simpler weighting matrices are preferable. Similarly, the choice of k_0 can be viewed as a tuning parameter for the procedure. This work was primarily concerned with rare event studies, however the role of this parameter can be more interesting for studies with a large number of cases. First, it would be cumbersome to include all the logistic estimators computed after each failure. Second, we need to cover the whole time interval $(0, \tau)$ rather than discard first k_0 estimators. It is therefore more natural to change the procedure so that k_0 is the fraction of logistic estimators that are used and to take these estimators systematically from the whole time interval.

Finally let us mention the issue of stratification. In principle there is no problem to incorporate stratification into the estimation procedure. Stratification in logistic regression models only affects the intercept – instead of one α we have a separate parameter for each stratum. The odds ratio parameters β_j remain unchanged. In practice we need to invert large matrices, which become even larger due to stratification. Therefore we only recommend using stratification with a small number of strata.

Bibliography

Barlow, W. (1994). Robust variance estimation for the case-cohort design. *Biometrics* 50 1064–1072.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Stat.Med.* 2 273–277.

Borgan, O., Langholz, B., Samuelsen, S. O., Goldstein, L. & Pogoda, J. (2000). Exposure stratified case-cohort designs. *LDA* 6 39–58.

Breslow, N. E. (2005). Case-control studies. In W. Ahrens & I. Pigeot, eds., *Handbook of epidemiology*. Springer, 287–319.

Broyden, G. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications* 6 76–90.

Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing* 16 1190–1208.

Chen, H. (2001). Fitting semiparametric transformation regression models to data from a modified case-cohort design. *Biometrika* 88 255–268.

Chen, K., Jin, Z. & Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* 89 659–668.

Chen, K. & Lo, S. (1999). Case-cohort and case-control analysis with cox's model. *Biometrika* 86 755–764.

Cheng, S., Wei, L. & Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* 82 835–845.

Cheng, S., Wei, L. & Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *JASA* 92 227–235.

Cox, D. R. (1972). Regression models and life tables (with discussion). *JRSS B* 34 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* 62 269–276.

Fine, J., Ying, Z. & Wei, L. (1998). On the linear transformation model for censored data. *Biometrika* 85 980–986.

Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John Wiley and Sons Inc.

Goldstein, L. & Langholz, B. (1992). Asymptotic theory for nested case-control sampling in the cox regression model. *AnnStat* 20 1903–1928.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Staistical Analysis of Failure Time Data, Second Edition*. New Jersey: John Wiley and Sons Inc.

Klášterecký, P. & Kulich, M. (2006). A note on parameter estimation in regression models for case-cohort data. In A. J. & G. Dohnal, eds., *Proceedings of Robust'06*. JČMF, Prague, 127–134.

Kong, L., Cai, J. & Sen, P. (2004). Weighted estimating equations for semiparametric transformation models with censored data from case-cohort design. *Biometrika* 91 305–319.

Kulich, M. & Lin, D. Y. (2000). Additive hazards regression with covariate measurement error. *JASA* 95 238–248.

Kulich, M. & Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *JASA* 99 832–844.

Kupper, L. L., McMichael, A. J. & Spirtas, R. (1975). A hybrid epidemiologic study design useful in estimating relative risk. *JASA* 70 524–528.

Lu, W. & Tsiatis, A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* 93 207–214.

Mantel, N. (1973). Synthetic retrospective studies and related topics. *Biometrics* 29 479–486.

McCullagh, P. (1980). Regression models for ordinal data, with discussion. *JRSS B* 42 109–142.

McCulloch, C. E. & Searle, S. R. (2001). *Generalized, Linear and Mixed Models*. New York: John Wiley and Sons Inc.

Miettinen, O. S. (1982). Design options in epidemiologic research: an update. *Scand. J. Work Environ. Health* 8 7–14.

Murphy, S. A., Rossini, A. J. & van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *JASA* 92 968–976.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 73 1–11.

Prentice, R. L. & Breslow, N. (1978). Retrospective studies and failure time models. *Biometrika* 65 153–158.

Prentice, R. L. & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* 66 401–411.

Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1994). Estimation of regresion coefficients when some regressors are not always observed. *JASA* 89 846–866.

Vogel, U., Laros, I., Jacobsen, N., Thomsen, B., Bak, H., Olsen, A., Bukowy, Z., Wallin, H., Overvad, K., Tjø nneland, A., Nexø, B. & Raaschou-Nielsen, O. (2004). Two regions in chromosome 19q13.2-3 are associated with risk of lung cancer. *Mutation Research* 546 65–74.

Wacholder, S., Gail, M., Pee, D. & Brookmeyer, R. (1989). Alternative variance and efficiency calculations for the case-cohort design. *Biometrika* 76 117–123.